Statistical Confluence Analysis by Means of Complete Regression Systems

By

Ragnar Frisch.

STATISTICAL CONFLUENCE ANALYSIS BY MEANS OF COMPLETE REGRESSION SYSTEMS I—III.

CONTENTS.

4

INTRODUCTION.

THE OBJECT OF CONFLUENCE ANALYSIS. THE DANGER OF IN-
CLUDING TOO MANY VARIATES IN A REGRESSION ANALYSIS.

In a paper "Correlation and Scatter..." in "Nordic Statisti-
cal Journal" 1928 I drew attention to the fact that in statistical
regression analysis there exists a great danger of obtaining
nonsensical results whenever one includes in one and the same
regression equation a set of variates that contain two, or more,
subsets which are already — taken by themselves — highly
intercorrelated. Suppose, for instance, that we have three sta-
tistical variates x_1, x_2, x_3 (measured from their means), and that
we know for apriori reasons that there exist not only one but
two independent linear equations between them (since the vari-
ates are measured from their means, we may assume the equa-
tions to be homogeneous). Further, suppose that a great number
of observations are made, each observation giving the values
of the three variates and being represented as a point in the
three dimensional (x_1, x_2, x_3) space. All these observation points
would then lie on a straight line through origine in (x_1, x_2, x_3)
space. From the distribution of these points it would be absurd
to try to determine the coefficients of any of the two equations
that we know apriori should exist between the variates. Indeed,
a set of points lying in *a line* does not contain enough informa-
tion to determine *a plane*. More precisely the coefficients of
this plane would contain a one dimensional indeterminateness.
In this situation any attempt at determining from the available
data a regression equation involving *three* variates would be
sheer nonsense.

If nevertheless such an attempt was made, the regression
coefficients would — if enough decimal places were carried in
the computation — turn out to be of the indeterminate $\frac{0}{0}$ form.
If errors of observation are present, the regression coefficients
would not be exactly of the form $\frac{0}{0}$, but would now appear in
the form of an error of observation divided by another error of
observation. On the face of it the result of the computation

would thus be determinate but it would still have no meaning as an expression for a regression coefficient. We would have *fictitious determinateness created by random errors*.

When several variates are included in the analysis, the situation becomes of course much more complex. We may here encounter a whole hierarchy where some of the variates may form a set where a regression equation has a meaning, and others forming sets where such equations have no meaning, The study of this hierarchy is what I call *confluence analysis*. It is an important part of statistical analysis, particularly in the social sciences. Indeed, the data will frequently obey many more relations than those which the statistician happens to think of when he makes a particular regression study. If the statistician does not dispose of an adequate technique for the statistical study of the confluence hierarchy, he will run the risk of adding more and more variates in the study until he gets a set that is in fact multiple collinear and where his attempt to determine a regression equation is therefore absurd.

In practice these cases are apt to arrive much more frequently than is usually recognised. As a matter of fact I believe that a substantial part of the regression and correlation analyses which have been made on economic data in recent years is nonsense for this very reason.

The sampling theory as we know it today does *not* furnish criteria that can distinguish between these various cases. Indeed, the standard errors on the regression coefficients, and most of the other parameters used in sampling theory *become themselves indeterminate* in those cases which it is here wanted to analyse, or more precisely expressed they get fictitious determinateness created by random errors. I indicated this theoretically in the above mentioned paper in Nordic Statistical Journal of 1928, and the example discussed in Section 33 of the present paper will illustrate clearly how inadequate the tools of sampling theory are for the study of the problems which we encounter in confluence analysis.

Of course, this contains no reflection on the value of sampling theory in general. In problems of the kind encountered when the data are the result of *experiments* which the investigator can control, the sampling theory may render very valuable services. Witness the eminent works of R. A. Fisher and Wishart on problems of agricultural experimentation.

In the 1928 paper "Correlation and Scatter..." I made a first attempt at developing statistical criteria for the various cases of confluence hierarchy. The criteria were based on a study of the smallness of *the correlation determinants* in the various subsets. The square roots of these determinants I termed scatter coefficients. In practice one will most frequently work with the determinant values themselves, which may perhaps be called scatterances [1].

The weak point in the method I suggested in 1928 was that no criteria were developed for judging the *significance* of the scatterances deviation from zero. In the subsequent years I reverted to the question on and off, on various occasions, attempting to push the analysis further. The line of approach which suggests itself from the view-point of sampling theory is to attempt to find the sampling distribution of the scatterances. I did not concentrate much on this aspect of the problem, primarily because I felt that — at least when the data are of an economic sort — this would not be the most fruitful way of approach. Indeed, if the sampling aspect of the problem should be studied from a sufficiently general set of assumptions, I found that it would lead to such complicated mathematics that I doubted whether anything useful would come out of it. And, on the other hand, if the sampling aspect should be studied under simple assumptions, for instance, of not collinear and normally distributed basic variates, the essence of the confluence problem would not be laid bare. One would then again get back to such a situation where those higher parameters (standard errors, etc.), by which the first set of parameters (scatterances, etc.) are to be judged, *themselves* become indeterminate in just those cases that interest us from the confluency view-point. One would have to consider standard errors of the standard errors, and standard errors of these higher standard errors again, and so on, up to such a high level of the standard error hierarchy, that the utility of this whole apparatus would become very dubious.

I decided therefore first to attack the problem more from the experimental side, working out numerically — on actual economic data as well as on constructed examples — various other

types of criteria which intuitively and heuristically may suggest themselves. These experiments converged towards a definite method which, after applications to various kinds of data, was found to give satisfactory and plausible results. The present paper gives an account of this work.

Part I describes some of the procedures of confluence analysis with which I experimented before reaching the more satisfactory method. While these tentative procedures described in Part I do not, taken by themselves, give final, conclusive criteria of confluency and linear significance, they are not wholly without interest because they exhibit the nature of the difficulties involved, and may — when used with care -- help making a rough preliminary analysis of the data.

Part II gives the theoretical background of the further analysis. It discusses the distinction between systematic variations and disturbances and from this draws certain conclusions regarding the "true" regression. In particular, there are pointed out certain facts regarding the connection between the "true" regression and the empirical results obtained by least square minimalisation in different directions. These facts give the leading ideas of the subsequent method. In this connection is also outlined a general scheme of interpretation for the various principles of determining linear regressions which are in common use or may suggest themselves as plausible.

Part III develops the method which I am now recommending as the most conclusive and which I feel gives a rather satisfactory solution of the main problems of linear confluence analysis. The computing technique to be used in this connection is described in Section 15 and the essence of the principles of interpretation are developed in Sections 16—18. Those who are primarily interested in results will probably find these four sections the most important in the present work.

Part IV gives numerical examples illustrating in detail the application of the technique proposed. One of these examples is based on constructed data, in order to give a means of checking how the tests proposed work in practice. This example also shows how utterly inadequate the usual sampling error analysis is as a means of testing significance when the data are linearly confluent. The other examples are drawn from actual data, particularly American consumption and sales statistics.

As a by-product in the study of these latter examples is obtained — by means of two different reference commodities (meat and butter) — a determination of the money flexibility that may be compared with the results which I found in 1930 by using U. S. budget data.

The present study has been undertaken as an indispensable preliminary step for certain projects, namely statistical productivity studies and statistical construction of econometric functions (demand and supply curve and the like) that are planned as part of the research programme of the University Institute of Economics, Oslo. It is to be hoped, however, that the results here presented may be applicable, not only to our special problems, but more generally to various kinds of problems where the statistical confluency of the data is an important feature to take into consideration.

PART I: CONFLUENCE ANALYSIS BY MEANS OF TEST-PARAMETERS.

1. CORRELATION COEFFICIENTS AND SCATTERANCES.

Most of the work in linear regression analysis can be based on cross moments and correlation coefficients. If $x_1 \ldots x_n$ are the observational variates measured from their means, and the Gaussian symbol [ ] denotes a summation over all the observations, the moments are

$$(1.1) \qquad\qquad m_{ij} = [x_i \, x_j]$$

and the ordinary (gross) correlation coefficients are

$$(1.2) \qquad\qquad r_{ij} = \frac{m_{ij}}{\sqrt{m_{ii} \, m_{jj}}}.$$

The standard deviations of the variates are

$$(1.3) \qquad\qquad \sigma_i = \sqrt{\frac{m_{ii}}{N}}$$

where N is the number of observations. The *standardised* variates are

$$(1.4) \qquad\qquad \xi_i = \frac{x_i}{\sigma_i}.$$

The standardised variates have unit standard deviation. If the variates are normalised in such a way that their *sumsquare* over all the observations (not their standard deviation) is unity, we get a set of variates whose cross moments are the same as their correlation coefficients. Of course the difference between the variates normalised in this way and the standardised variates only lies in the factor $\sqrt{N}$.

I am here taking the correlation coefficients (1.2) simply as a set of computational parameters without making any attempt at interpreting them as a measure of the strength of the relation between the variates.

In practical work it will be found more convenient to handle the correlation coefficients than the moments, particularly because the former are reduced to a common order of magnitude with the same meaning of the decimal places in all the magnitudes treated.

In the computation of moments and correlation coefficients we have at the Oslo Institute developed certain checks which we have found very useful, but which are not, as far as I know, commonly practised. It may therefore be worth while to mention them in the present connection. We introduce — as in the usual checking technique — the sum variate

$$(1.5) \qquad x_0 = x_1 + \ldots + x_n.$$

But instead of computing all the crossmoments m_{i0} and verifying for each i that

$$(1.6) \qquad m_{i0} = m_{i1} + \ldots + m_{in}$$

we simply take the sumsquare m_{00} and verify that

$$(1.7) \qquad m_{00} = m_{11} + \ldots + m_{nn} + 2 \sum_{i<j} m_{ij}$$

where the summation $\sum_{i<j}$ means the sum of all the $\binom{n}{2}$ crossmoments that express interconnections between different variates.

If it is wanted to apply a check that can locate an error to a smaller section of the work, we rather prefer to split the series of observations into *ranges*, say of 10 or 20 observations in each range, and then apply (1.7) to each range.

Generally we form these range-moments before reducing the variates to measurement from their means. The only thing we do to the raw data before computing the moments is to add zeros to some of the variates in order to make all of them, roughly speaking, of the same order of magnitude, and sometimes to subtract a provisoric mean in round numbers if the figures are such that this can be done very quickly. From these modified variates we form the sum variate (1.5), and compute its sum S_0 over all the observations. This sum is checked against the sums S_i of the individual variates. (If

necessary this checking may also be split into ranges.) A listing adding machine is convenient for this purpose.

The origine moments of the variates modified by the above procedures may be denoted

$$(1.8) \qquad\qquad M_{ij}.$$

These are the crossmoments that are formed and checked for each range by (1.7). The total moments derived from the range moments are also checked by (1.7). This being done, all the rest of the work is only concerned with total results, not with range results.

The "multiplied" mean moments

$$(1.9) \qquad\qquad s_{ij} = N \cdot M_{ij} - S_i S_j$$

are formed (each quantity s_{ij} being computed in one operation on the machine). The results are checked by computing the row sums

$$(1.10) \qquad\qquad s_{i0} = s_{i1} + \ldots + s_{in}$$

and verifying that the sum of these row sums is equal to s_{i0} computed directly by (1.9). This check is equivalent to applying (1.7); of course (1.7) also holds for the s_{ij}.

It will be noticed that (1.7) involves much less extra multiplication than (1.6), and (1.7) is just as safe, with one exception, namely that (1.7) does not register a mistake in *interchanging* two (or more) of the moments. Particular attention should therefore be given to the correct location of the figures in the tables. A good safeguard against mistakes of this sort is to compute each crossmoment, for instance M_{12}, through for all ranges before a new crossmoment is taken. This means that all checking by (1.7) is left to the end of the moment work, and then all range-moments recomputed that does not check right away.

In terms of the s_{ij} the correlation coefficients are

$$(1.11) \qquad\qquad r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}}$$

The dividing out by the factors in the numerator of (1.2)

or (1. 11) is usually — so far as a systematic check is concern-
ed — a weak point in correlation computations. We check this
part of the work as follows. First the numbers $\sqrt{s_{ii}}$ and $1/\sqrt{s_{ii}}$
$(i = 1, 2 \ldots n)$ are computed and checked individually by squaring
and multiplication. Then the correlation coefficients are com-
puted by (1. 11) and the row sums

$$(1. 12) \qquad\qquad r_{i0} = r_{i1} + \ldots + r_{in}$$

are formed. By means of these it is checked that

$$(1. 13) \qquad\qquad \sum_i \sqrt{s_{ii}} \cdot r_{i0} = \sum_i \frac{1}{\sqrt{s_{ii}}} \cdot s_{i0}$$

where s_{i0} are the row sums (1. 10) previously computed.

If (1. 13) does not check right away, each row in the corre-
lation matrix may be checked separately by

$$(1. 14) \qquad\qquad \sqrt{s_{ii}} \cdot r_{i0} = \sum_j \frac{1}{\sqrt{s_{jj}}} \cdot s_{ij}.$$

The checks (1. 13) and (1. 14) contain of course also a verifica-
tion of the rootsquaring of the s_{ii} and on the divisions
$1/\sqrt{s_{ii}}$, so that actually all steps are checked by this method. In
practical work with several variates this checking technique
has been found very helpful.

In all the matrices here considered M_{ij}, s_{ij}, r_{ij} etc. only the
diagonal and one of the triangles need to be filled in, since the
matrices are symmetric. As a rule we use the north-east
triangle. Taking a row sum (or a column sum) in such a
matrix means taking the sum of the elements in a broken line
reflected under 45° on the principal diagonal.

The correlation determinants, i. e. the scatterances, we denote

$$(1. 15) \qquad\qquad \triangle = \begin{vmatrix} r_{11}\, r_{12} \ldots r_{1n} \\ \cdots\cdots\cdots \\ r_{n1}\, r_{n2} \ldots r_{nn} \end{vmatrix}$$

In the subset $(ij \ldots k)$ we use the notation

$$(1.16) \qquad \triangle_{ij\ldots k} = \begin{vmatrix} r_{ii}\, r_{ij} \ldots r_{ik} \\ r_{ji}\, r_{jj} \ldots r_{jk} \\ \cdots\cdots\cdots\cdots \\ r_{ki}\, r_{kj} \ldots r_{kk} \end{vmatrix}$$

These determinants in all possible subsets are computed most easily by the tilling technique described in Section 15.

The diagonal elements in (1.15) and (1.16) are all equal to unity. If these diagonal elements are replaced by zeros, we get the "hollow" correlation determinants.

$$(1.17) \qquad \Gamma = \begin{vmatrix} 0\; r_{12} \ldots r_{1n} \\ \cdots\cdots\cdots\cdots \\ r_{n1}\, r_{n2} \ldots 0 \end{vmatrix} \quad \text{and similarly } \Gamma_{ij\ldots k}.$$

These hollow determinants are convenient for certain computation purposes.

For very large n an approximation to the scatterance $\triangle$ may be computed by retaining only the first terms in the expansion (2.9) and computing by the following formulae the first of the hollow determinats Γ from which the B's (mentioned in section 2) are built up:

If we write the expressions for Γ out explicitely we find that the first few even-rowed determinants of this type can be expressed by means of the following two sets of magnitudes which themselves can be computed recurrently. We first define

$$(1.18) \qquad r_{ijkh} = \sum_{\alpha < \beta} r_{\alpha\beta}\, r_{rem}$$

where the summation runs through combination without repetition of the two affixes α, β selected amongst the four affixes $(ijkh)$, r_{rem} means the correlation coefficient that has the two affixes which *remain* when α and β are taken away from the set $(ijkh)$. Writing (1.18) out in full for (1234) we get for instance

$$(1.19) \qquad r_{1234} = r_{12}\cdot r_{34} + r_{13}\cdot r_{24} + r_{23}\cdot r_{14} + r_{14}\cdot r_{23} + r_{24}\cdot r_{13} + r_{34}\cdot r_{12},$$

that is

$$(1.20) \qquad r_{1234} = 2\cdot(r_{12}\cdot r_{34} + r_{13}\cdot r_{24} + r_{23}\cdot r_{14}).$$

Similarly we define

$$(1.21) \qquad r_{ijkhlm} = \sum_{\alpha < \beta} r_{\alpha\beta} \cdot r_{rem\text{-}four}$$

where $r_{rem\text{-}four}$ is the quantity r with the four affixes that remain when (α, β) are taken away from the set $(ijkhlm)$. In this way we may continue and define r with any even number of subscripts.

A similar system may be built on the Γ's. Here we need however a more elaborate classification, namely one that indicates the various ways in which the affixes on Γ may be divided into *subgroups*. The number of affixes in each subgroup we shall indicate by superscripts. For the hollow determinants themselves we write

$$(1.22) \qquad \Gamma_{ij}^{(2)} = \Gamma_{ij}, \qquad \Gamma_{ijk}^{(3)} = \Gamma_{ijk} \text{ etc.}$$

Then we define

$$(1.23) \qquad \Gamma_{ijkh}^{(2\cdot2)} = \sum_{\alpha < \beta} \Gamma_{\alpha\beta}^{(2)} \cdot \Gamma_{rem}^{(2)}$$

(α, β) running as before through combinations without repetition in the set $(ijkh)$. Writing out $\Gamma_{1234}^{(2.2)}$ in full we get a formula similar to (1.20), only with Γ instead of r. Further we define

$$(1.24) \qquad \Gamma_{ijkhlm}^{(2\cdot2\cdot2)} = \sum_{\alpha < \beta} \Gamma_{\alpha\beta}^{(2)} \; \Gamma_{rem}^{(2\cdot2)}$$

which is analogous to (1.21). On the other hand

$$(1.25) \qquad \Gamma_{ijkhlm}^{(2\cdot4)} = \sum_{\alpha < \beta} \Gamma_{\alpha\beta}^{(2)} \; \Gamma_{rem}^{(4)}$$

$$(1.26) \qquad \Gamma_{ijkhlm}^{(3\cdot3)} = \sum_{\alpha < \beta < \gamma} \Gamma_{\alpha\beta\gamma}^{(3)} \; \Gamma_{rem}^{(3)} \text{ etc.}$$

are new types of combinations not represented in the r's. In (1.25) we would of course have obtained the same expression by extending the summation to combinations without repetition of the 4 affixes $(\alpha\beta\gamma\delta)$ and letting the remaining affixes be only two in number.

By means of the above symbols the first few evenrowed Γ's may be expressed as follows

$$(1.27) \qquad \Gamma_{ij} = -r_{ij}^2.$$

(1.28) $$\Gamma_{ijkh} = \Gamma_{ijkh}^{(2\cdot2)} - \left(\tfrac{1}{2\,!}\,r_{ijkh}\right)^2$$

(1.29) $$\Gamma_{ijkhlm} = 2\,\Gamma_{ijkhlm}^{(2\cdot4)} + \tfrac{1}{2}\,\Gamma_{ijkhlm}^{(3\cdot3)} - \Gamma_{ijkhlm}^{(2\cdot2\cdot2)} - \left(\tfrac{1}{3\,!}\,r_{ijkhlm}\right)^2$$

In order to express the un-even-rowed Γ's we need the magnitudes $\Gamma_{ijkhl}^{(2\cdot3)}$ etc. defined similarly to (1.25) and (1.26), and further these magnitudes defined for an *incomplete summation*, that is by letting the summation run over all affixes except some specificed affix p. This leads to defining

(1.30) $$\Gamma_{p\,\cdot\,ijkhl}^{(2\cdot3)} = \sum_{\alpha<\beta}{}^{p(}\,\Gamma_{\alpha\beta}^{\cdot(2)}\,\Gamma_{rem}^{(3)}$$

where the summation runs over combinations without repetition of the two affixes $\alpha\beta$ picked amongst the set of four affixes obtained by leaving p out of the set $(ijkhl)$. Similarly we define the incomplete r

(1.31) $$r_{p\,\cdot\,ijkhl} = \sum_{\alpha}{}^{p(}\,r_{p\alpha}r_{rem\text{-}four}$$

where α runs through $(ijkhl)$ except p; p may be called the "skew affix" in (1.30) and (1.31). As an example we may take

(1.32) $$r_{1\cdot2345} = r_{12}\cdot r_{1345} + r_{13}\cdot r_{1245} + r_{14}\cdot r_{1235} + r_{15}\cdot r_{1234}.$$

In terms of the quantities (1.30) and (1.31) we have

(1.33) $$\Gamma_{ijk} = 2\cdot r_{ij}\cdot r_{ik}\cdot r_{jk}$$

(1.34) $$\Gamma_{ijkhl} = \Gamma_{ijkhl}^{(2\cdot3)} + \Gamma_{i\,\cdot\,ijkhl}^{(2\cdot3)} + \tfrac{1}{4}\,r_{i\,\cdot\,ijkhl}\cdot r_{jkhl}.$$

The formula (1.34) is developed with i as the "skew" affix, but we may just as well use any of the other affixes for this purpose.

The above formulae have little interest when it is wanted to compute all possible scatterances in a set of reasonably large dimensionality. In that case one will use the tilling technique of Section 15. But the above explicit Γ-formulae may serve to determine successive approximations to a given scatterance of very high order. Indeed the successive terms $B_0,\,B_1,\,B_2$ etc. of the formula (2.9) of the next Section may be locked upon as successive terms in a Taylor series for $\triangle$. B_0 is equal to 1, B_1

equal to zero, B_2 the sum of all two-rowed Γ in the big set for which the scatterance is to be computed, etc. By letting (ij) in (1. 27), (ijk) in (1. 33) etc. run through all possible combinations in the big set we thus obtain the first terms of the scatterance in question.

As an example of the nature of the approximation obtained by these successive terms we take the following scatterance in the potato data of Section 31.

TABLE (1. 35). POTATO DATA. SUCCESSIVE TERMS IN $\triangle_{12345678}$.

First term	B_0	$=$	1.000000
Second term	B_2	$= -$	1.793000
Third term	B_3	$=$	1.448480
Fourth term	B_4	$= -$	.392633
Fifth term	B_5	$=$	.010346
Sixth term	B_6	$=$	.010648
Seventh term	B_7	$= -$	.001132
Eight term	B_8	$= -$	.000005
Total $\triangle$ (in eight set) $= \Sigma_i\ B_i$		$=$	.282704

How can the scatterances be interpreted as indicators of linear confluency?

In the first place we note that the general tendency of the $\triangle$-s will be to be all the smaller, the better the linear connection in the set considered. (For a geometric interpretation of the scatterances from this view-point see Section 2 in Part II of "Correlation and Scatter ..."). But we are not only concerned with finding a set about which it can be said that its variates are linearly connected. From the confluency view-point it is just as important to ascertain that the set considered is *simply* collinear, which means that not all its first subsets are collinear. (The first subsets are the sets obtained by leaving out one of the variates at a time). Therefore some comparison must be made between the scatterance in the set considered and the scatterances in the various first subsets. These latter we shall call the *subscatterances*.

The general tendency which we must first of all look for when we compare subscatterances and the scatterance in the bigger set is whether there is a *sharp decline* when we pass from the former to the latter. We are particularly interested in seeing whether the scatterance in the bigger set is much

smaller than the *smallest* of the subscatterances. If this is not so, the new variate that is added as we pass to the bigger set cannot be looked upon as important. But on the other hand it does not do any great harm. It is rather a neutral variate. For the moment we shall not go into any more detailed discussion of the nature of the variate in this case. The complete analysis in this case cannot be given only on the basis of the scatterances, and must therefore be postponed till later (see in particular Section 17).

Only if there is a substantial drop in the scatterances do we have a situation where the passage to the bigger set *may* constitute a significant progress in the analysis. But here we must be careful: there are two possibilities. A sharp drop in the observed scatterances will be produced, not only if there exists in the bigger set *one* linear connection between the variates which are so much more perfectly fulfilled than a linear connection in the subsets, but a sharp decline may be produced also *if all the subsets were already systematically connected.* Indeed, in this case the *erratic* element will get a much smaller chance of keeping the scatterance in the bigger set up from zero. This is plausible already intuitively, and can also be deduced from the theoretical considerations of Section 8 and the numerical examples in Part IV. Thus a sharp decline as we pass from the subscatterances to the scatterance in the bigger set may be either a warning signal that we get into a multiply collinear set where a regression equation has no meaning, or it may be a criterion that we get a set where the regression equation is more exact than before.

Incidentally this shows how absurd it is to use the multiple correlation coefficient in the way in which it is usually employed. The multiple correlation coefficient is indeed essentially determined by the ratio between a scatterance and a subscatterance. (The reader may for instance compute the multiple correlation coefficient of 4 on (123) in the meaningless set (1234) in Section 23. It turns out 0.99. The scatterances needed for this computation are to be found in the tilling tables of Section 23).

Which one of the above two alternatives we are confronting depend on whether the deviations from zero of the subscatterances were systematic or accidental. Is there any feature in the *distribution* of the subscatterances that can reveal anything

about this? If we are to reckon with all conceivable possibilites, then virtually any observed distribution of the scatterances can be interpreted just as well as due to erratic influences as due to systematic ones. But practically speaking it seems probable that there is all the more chance for obtaining an *even* distribution of the subscatterances the more exclusively they are determined by random errors. Therefore, if one or more of the subscatterances deviate considerably from the minimum subscatterance, it seems plausible to conclude *that we do not have multiple collinearity in the bigger set.* But when all the subscatterances are more or less equal we must be prepared for such a possibility. Of course it may conceivably happen that the subscatterances turn out approximately equal even if their deviation from zero are essentially systematic, but so long as we base our confluency conclusions only on scatterances we have no means of recognising this case. If we want to play safely we ought therefore to refuse to pass on to the bigger set whenever the subscatterances are nearly equal and of such a size that it is not obvious that they are systematically different from zero. The method of Part III will furnish a more refined criterion that permits a more definite conclusion even in the case which we must thus leave in suspense when we use only scatterances.

To resume we may formulate the rule for the interpretation of scatterances as follows:

I.   Let n variates be observed. If it is contemplated to determine a regression equation containing ν of these variates, all possible ν-dimensional sets which can be formed in the big n-dimensional set should be investigated and that one, or those ν-dimensional sets which have the smallest scatterance should be selected for a further scrutiny.

II.   In each such ν-dimensional set selected for further scrutiny all the subscatterances should be considered. If the scatterance in the ν-dimensional set is not appreciably smaller than the smallest subscatterance contained in it, there is neither any great harm nor any great use in considering the ν-dimensional set in question. We could just as well be satisfied with that $(\nu-1)$-dimensional set which has the smallest subscatterance.

III. If there is a sharp decline as we pass from the subscatter-

ances to the scatterance in the ν-dimensional set, two possibilities are present. Either this decline means that it is a significant progress to form a regression equation in this ν-dimensional set, or it means that it is particularly dangerous. If at least one of the subscatterances are great[1], it is probably safe to pass on to the bigger set. Even if all the subscatterances are small it may be advisable to accept the bigger set, provided that there is a considerable spread in the subscatterances, for instance if there is one subscatterance that is decidedly smaller than the other subscatterances or if there is at least one subscatterance that is definitely larger than the others.

IV. But if all the subscatterances are about equal, and rather small, then the bigger set must be refused even though its scatterance is much less than the subscatterances. Conceivably it might even in this case have been correct to accept the bigger set, but the scatterances do not give a means of finding out whether this is permitted or not.

V. If by the above criteria there are more than one ν-dimensional set which it seems plausible to accept, one should proceed to scrutinising all the $(\nu + 1)$-dimensional sets according to the criteria (I) — (IV), *omitting however any set that contains a subset which by (IV) has already been recognised as dangerous.*

2. THE ELLIPSOID METHOD AND THE CHARACTERISTIC POLYNOMIAL.

When I started the experimental work of trying to find other criteria that may replace or supplement the scatterances as tests of confluency, the first idea followed up was to study the characteristic roots of the correlation determinants. The characteristic roots $\lambda_1 \ldots \lambda_n$ of the determinant (1.15) are defined as the zeros of the characteristic polynomial

$$(2.1) \qquad P(\lambda) = \begin{vmatrix} r_{11} - \lambda & r_{12} \ldots & r_{1n} \\ r_{21} & r_{22} - \lambda \ldots r_{2n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ r_{n1} & r_{n2} \ldots & r_{nn} - \lambda \end{vmatrix}$$

[1] For instance, not very far from being of the same order of magnitude as that which would be expected in a scatterance for the same number of random variates. Some information about this order of magnitude may be obtained from the tables of Section 30.

The expansion of this polynomial is

$$(2.2) \qquad P(\lambda) = A_n - A_{n-1}\lambda + A_{n-2}\lambda^2 + \ldots + (-)^n A_0 \lambda^n$$

where A_k is the sum of all the $\binom{n}{k}$ k-rowed principal minors in (1.15). By convention $A_0 = 1$. Obviously A_n is nothing but the determinant (1.15) itself. Introducing $1 - \lambda = \zeta$, the expansion takes on the form

$$(2.3) \qquad P(\lambda) = Q(\zeta) = B_n + B_{n-1}\zeta + \ldots + B_0 \zeta^n$$

where the B's are the corresponding sums of principal minors in the hollow determinant.

In the expansion (2.3) the second highest power of ζ is lacking because the one-rowed principal minors in the hollow determinant (1.17) consist only of zeros.

The value of the characteristic polynomial for a given value of λ (or of ζ) may be computed directly by inserting the value of λ (or ζ) in (2.1) and evaluating the determinant as it stands. Or the computation may be made from (2.2) or (2.3) by first evaluating the coefficients A or B. The latter method is to be preferred when many ordinates are wanted.

Since all the characteristic roots are real, there is in general no particular difficulty in determining them by one of the usual approximation methods. In a three dimensional set they are easily determined in explicit form from $Q(\zeta)$ since the second highest power of ζ is here lacking.

The individual terms of the coefficients A — that is the scatterances themselves — are determined most easily by the tilling technique of Section 15. If it is wanted to determine the coefficients A in all subsets $(\alpha, \beta \ldots \gamma)$ they may be built up from the scatterances by the recurrence formula

$$(2.4) \qquad (\nu - k)A_{k(\alpha, \beta \ldots \gamma)} = \sum_{i = \alpha, \beta \ldots \gamma} A_{k(\alpha, \beta \ldots)i(\ldots \gamma)}$$

Where $(\alpha, \beta \ldots \gamma)$ is any ν-dimensional set, and $A_{k(\alpha, \beta \ldots \gamma)}$ the coefficient of $(-\lambda)^k$ in this set. The inverted parenthesis denotes "exclusion of". For instance for $\nu = 4$, $k = 2$, $2A_{2(1234)} = A_{2(123)} + A_{2(124)} + A_{2(134)} + A_{2(234)}$. As a convenient check on (2.4) we have the fact that the magnitude determined as the sum of the elements in the right member of (2.4) shall be divisible by $(\nu - k)$.

The recurrent computation of the A's by (2.4) can be checked for each ν-level by the formula

$$(2.5) \qquad \sum_{\alpha < \beta < \ldots \gamma} A_{k(\alpha\beta\ldots\gamma)} = \binom{n-k}{\nu-k} \cdot A_{k(12\ldots n)}.$$

The sum in the left member of this formula is simply the sum of all the A_k in all possible ν-sets contained in the big set $(12\ldots n)$.

The formula (2.5) is obtained by the following consideration. The sum in the left member of (2.5) obviously contain all possible k-rowed $\triangle$ contained in the big set, each such k-rowed $\triangle$ being involved a certain number of times. In other words the left member of (2.5) must be equal to $A_{k(12\ldots n)}$ multiplied by a certain integer. This integer is determined simply by comparing the total number of terms in the left member of (2.5) with the total number of terms in $A_{k(12\ldots n)}$. The former number is equal to the number of terms in each $A_{k(\alpha\beta\ldots\gamma)}$ namely $\binom{\nu}{k}$ times the number of $A_{k(\alpha\beta\ldots\gamma)}$ entering into the summation in the left member of (2.5), namely $\binom{n}{\nu}$. And the number of terms in $A_{k(12\ldots n)}$ is $\binom{n}{k}$. The integer in question is consequently equal to

$$\binom{n}{\nu}\binom{\nu}{k} \bigg/ \binom{n}{k} = \binom{n-k}{\nu-k}.$$

A similar formula obviously holds good for the $B_{k(\alpha\beta\ldots\gamma)}$ or more generally for any set of magnitudes that are built up in a similar way as a sum of elements defined for each $(\alpha\beta\ldots\gamma)$ combination.

In practice the check will take the form that on each ν-level, all the numbers in the left member of (2.4) are computed. The $A_{k(12\ldots n)}$ are formed directly, and it is verified that (2.5) holds good.

From the A's we may pass to the B's or vice versa by the formula

$$(2.6) \qquad B_{k(\alpha,\,\beta\ldots\gamma)} = \sum_{h=0}^{k} (-)^{h+k} \binom{\nu-h}{k-h} A_{h(\alpha,\,\beta\ldots\gamma)}$$

and

$$(2.7) \qquad A_{k(\alpha,\,\beta\,\ldots\,\gamma)} = \sum_{h=0}^{k} \binom{\nu-h}{k-h} B_{h(\alpha,\,\beta\,\ldots\,\gamma)}.$$

For $k = \nu$ (2.6) gives in particular in the big set, i. e. for $\nu = n$

$$(2.8) \qquad \Gamma = \sum_{h=0}^{n} (-)^{n+h} A_h.$$

From (2.7) we similarly get

$$(2.9) \qquad \triangle = \sum_{h=0}^{n} B_h.$$

As a final check on all the quantities A the "hollow" determinant $B_n = \Gamma$ in the big set may be computed both by (2.8) and by evaluating the determinant directly. The formulae (2.4) to (2.7) may also be utilized for various other checking purposes; for instance if the B's in the big set have been computed, but not the A's, and if one proceeds to computing the A's, in all subsets, then on each ν-level (2.5) may be used as a check by inserting in its right member the expression for the big set A's in terms of the big set B's.

When the coefficients A or B are determined, the values of the polynomial $P(\lambda)$ or — which amounts to the same — of $Q(\zeta)$ are most easily computed by ordering the terms according to the principle of "Chinese boxes", as follows:

$$(2.10) \qquad Q(\zeta) = \overline{\left|\,\overline{\left|\,\overline{\left|\,\overline{\zeta^2 + B_2}\,\right|\zeta + B_3}\,\right|\zeta + B_4}\,\right|\zeta + B_5 + \text{etc.}}$$

If a computing machine is available, which has an arrangement for transferring mechanically the figure standing in the "result", to the key-board, the computations by (2.10) can be done very quickly.

It is a classical fact that the characteristic roots are proportional to the square lengths of the main axes of the normal regression ellipsoids fitted to the observed scatter. This is easily seen as follows. The regression ellipsoids are defined by $\sum_{ij} \dot{r}_{ij}\, \xi_i\, \xi_j = C$ where the ξ_i are the scatter diagram coordinates, $\dot{r}_{ij}$ the elements of the reciprocal of the correlation matrix, and C a parameter which is constant along a given ellipsoid in the family. Let $c_1 \ldots c_n$ be the direction numbers for a given straight line trough origin. The equation of this line may be written

$\xi_i = Kc_i$ where K is a factor of proportionality. Determining K so as to get intersection with the ellipsoid we find $K^2 = C/\sum_{ij} \dot{r}_{ij} c_i c_j$. The square distance from origine to this point is consequently $\sum_i \xi_i^2 = K^2 \sum_i c_i^2 = \lambda C$ where

(2. 11) $$\lambda = \sum_i c_i^2/\sum_{ij} \dot{r}_{ij} c_i c_j.$$

Since a main axis of the ellipsoid is defined as a straight line from origin in a direction that makes its length measured from origin to the ellipsoid a minimum (for the short axes) or a maximum (for the long axes), the problem is to seek the extremum of λ. By partial derivation in the usual way this leads to the equation $\sum_j (r_{ij} - \lambda e_{ij}) c_j = 0$ for the c_j, where $e_{ij} = \begin{cases} 0 \text{ when } i \neq j \\ 1 \text{ when } i = j \end{cases}$. In order that this system shall have a solution (other than the trivial $c_i = 0$) it is necessary and sufficient that λ is a zero of the polynomial $P(\lambda)$ defined by (2. 1). Hence the characteristic roots of the matrix (r_{ij}) must be proportional to the square lengths of the main axes of the regression ellipsoids.

It therefore seems plausible to conclude that the observed scatter is systematically $\varkappa$-fold collinear when $\varkappa$ of the characteristic roots are "very small". This method may for shortness be called *the ellipsoid method*.

The main axes of the regression ellipsoid also have — as is well known — an immediate connection with the reduction of the variates to an uncorrelated form. Indeed, it is always possible to find a homogeneous non-singular linear transformation,

(2. 12) $$z_k = \sum_j c_{kj} \xi_j$$

such that the variates z_k become uncorrelated. The moments of the new variates will be

(2. 13) $$[z_h z_k] = \sum_{ij} r_{ij} c_{hi} c_{kj}$$

r_{ij} being the moments of the old variates, which is the same as their correlation coefficients, when the ξ_i are normalised in such a way that their sumsquares are unity. The problem is thus only to reduce a quadratic form (namely the one built over r_{ij}) to a sum of squares, and it is a classical fact that this is always possible. It is even possible to do it in an infinity of

ways (one method of carrying the computations through by a recurrence formula was indicated in Section 4 of my 1928 paper "Correlation and Scatter..."). But if we impose the supplementary condition that the transformation shall be *orthogonal* then it is uniquely determined (when the observed correlation matrix r_{ij} is nonsingular and the characteristic roots different). Indeed, the condition of non-correlation for the variates z_k is expressed by

$$(2.14) \qquad \sum_{ij} r_{ij} c_{hi} c_{kj} = \lambda_k e_{hk}$$

λ_k being certain numbers that it is not necessary to specify for the moment. If we insert in the right member of (2.14) the condition of orthogonality, namely

$$(2.15) \qquad \sum_j c_{hj} c_{kj} = e_{hk}$$

and notice that c_{hj} may be written $\sum_i e_{ij} c_{hi}$, (2.14) takes the form

$$(2.16) \qquad \sum_{ij} (r_{ij} - \lambda_k e_{ij}) c_{hi} c_{kj} = 0.$$

Looking upon this as a system of equations of the form $\sum_i c_{hi} U_i = 0 \, (h = 1, 2 \ldots n)$ where (c_{hi}) is a non-singular matrix, we see that we must have

$$(2.17) \qquad \sum_j (r_{ij} - \lambda_k e_{ij}) c_{kj} = 0 \qquad (i = 1, 2 \ldots n).$$

This shows that the λ_k are nothing but the characteristic numbers of the correlation-matrix and c_{kj} the direction numbers for the *k-th* main axis of the regression ellipsoid. In other words the transformed uncorrelated variates z_k are nothing but the coordinates of the scatter points measured along the main axes of the distribution ellipsoid.

By (2.12) the variates z are expressed in terms of the ξ. Inversely the ξ are expressed in terms of the z by the formula

$$(2.18) \qquad \xi_i = \sum_k z_k c_{ki}.$$

This follows from the fact that the reciprocal of an orthogonal matrix is simply its transposed. It should be remembered that the coefficients c_{ki} in (2.18) are by (2.15) normalised in such a way that

(2. 19) $$\sum_i c_{ki}^2 = 1 \qquad (k = 1, 2 \ldots n).$$

If we accept the above mentioned criterion that the observed scatter is systematically $\varkappa$-fold collinear when $\varkappa$ of the characteristic roots are "very small", then we would retain only the $n - \varkappa$ terms in the right member of (2. 18) which correspond to the $n - \varkappa$ characteristic roots judged to be significantly different from zero. For instance, if only one root namely λ_1 is deemed significantly different from zero, we would put

2. 20) $$\bar\xi_i = z_1 \cdot c_{1i}$$

where c_{1i} are the direction numbers determined from (2. 17) (for $k = 1$), and normalised according to (2. 19). If two roots λ_1 and λ_2 are significantly different from zero and significantly different from one another, we would put

(2. 21) $$\bar\xi_i = z_1 \cdot c_{1i} + z_2 \cdot c_{2i}$$
$$\text{etc.}$$

Theoretically this method seems very promising, but in practice I have found that, at least so for as the study of the unfolding capacity of the scatter in various kinds of economic data is concerned, it does not lead to any more conclusive results than the scatterance method. In studying the set of characteristic roots for a given correlation matrix we are indeed confronted with just the same kind of difficulty as in the study of scatterances. The big question also here is: Are the parameters considered *systematically* different from zero, or are they pushed away from zero just by the disturbances of the data? The usual sampling errors do *not* answer this question. And in the ellipsoid method we have in addition the question of whether the roots considered are systematically *different from each other*. It will presently appear that this latter question may in practice be particularly troublesome.

As an example of the difficulties inherent in the ellipsoid method I shall mention a study of an eight-variate problem which was undertaken jointly by Dr. Frederick V. Waugh, of the U. S. Department of Agriculture, and me during Dr. Waugh's work at the Institute in Oslo and in my seminar in

1932. I discribed this work in my lectures at the Institut Henri Poincaré in Paris in the Spring of 1933.

The object of the study was to see how the price obtained was influenced by the quality of potatoes in a certain New England market. The variates studied were:

1. Wholesale price.
2. Percentage of potatoes having a size between 1 ¾" — 2 ¼".
3. Percentage of misshapen potatoes.
4. Index of colour of the skin.
5. Index of bruises.
6. Cuts.
7. Scab.
8. Other features.

In this problem the questions of confluency is of course of paramount importance. We may for instance ask: Will the variate No. 5: Bruises represent an independent contribution towards the determination of the price, or is its influence already taken account of through some of (or all of) the other variates?

The correlation coefficients were as indicated in Table (2. 22). For the purpose of the subsequent experimental computation the correlations are here given with six decimal places, but only two or at the utmost three places are statistically significant.

TABLE (2. 22). POTATO DATA. GROSS CORRELATION COEFFICIENTS:

r_{ij}	$j=1$	2	3	4	5	6	7	8
$i=1$	1.000000	−0.210802	−0.215275	−0.452659	−0.315211	−0.198522	−0.228210	−0.279094
2		1.000000	0.102185	0.014595	−0.054787	0.081614	0.134882	0.094089
3			1.000000	0.264599	0.269739	0.214237	0.206190	0.309791
4				1.000000	0.443629	0.163629	0.254723	0.303490
5				.	1.000000	0.072961	0.198672	0.299576
6						1.000000	0.205694	0.407260
7							1.000000	0.361637
8								1.000000

Instead of computing just the eight values representing the characteristic roots we studied the whole shape of the

polynominal $P(\lambda)$ defined by (2. 1). This gives a much clearer view of the situation. The curve $P(\lambda)$ may then be looked upon as a sort of *spectrum* for the matrix. (2. 23) summarises the values of $P(\lambda)$ which were computed during the work.

Drawn on a reasonably large scale the curve $P(\lambda)$ appears as in Figure 1.
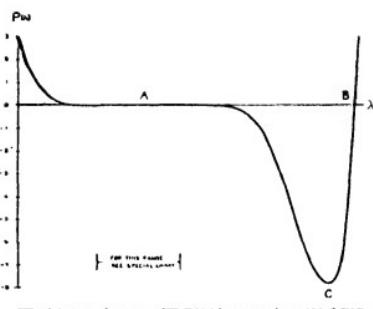
Fig. 1.

TABLE (2. 23). VALUES OF THE CHARACTERISTIC POLYNOMIAL FOR THE POTATO DATA.

λ	$P(\lambda)$	λ	$P(\lambda)$
0.000	0.28270786	0.902	$-$ 0.00000014
0.1	0.09045032	1.0	$-$ 0.00000547
0.2	0.02276385	1.01	0.00000690
0.3	0.00384628	1.1	0.00014828
0.4	0.00027373	1.15	0.00003985
0.475	0.00000271	1.155	0.00000629
0.48	0.00000101	1.16	$-$ 0.00003312
0.484	0.00000016	1.17	$-$ 0.00013244
0.485	0.00000001	1.2	$-$ 0.00063999
0.486	$-$ 0.00000012	1.3	$-$ 0.00692878
0.49	$-$ 0.00000045	2.0	$-$ 2.63268630
0.5	$-$ 0.00000021	2.3	$-$ 6.99993462
0.502	$-$ 0.00000004	2.6	$-$ 4.91675881
0.503	0.00000007	2.65	$-$ 1.88389024
0.55	0.00000508	2.673	$-$ 0.02001710
0.57	0.00000431	2.675	0.15796210
0.59	0.00000163	2.7	2.61512460
0.6	$-$ 0.00000018		
0.61	$-$ 0.00000210		
0.7	$-$ 0.00000938		
0.75	$-$ 0.00000293		
0.76	$-$ 0.00000166		
0.77	$-$ 0.00000068		
0.78	0.00000000		
0.79	0.00000021		
0.798	0.00000001		
0.8	$-$ 0.00000001		
0.9	$-$ 0.00002836		

The zero at B is obviously quite significant. Its exact value is $\lambda_1 = 2.6732264$, which gives the following direction numbers for the corresponding main axis

(2. 24)

$$c_{11} = -0.398108$$
$$c_{12} = -0.129359$$
$$c_{13} = 0.345343$$
$$c_{14} = -0.420703$$
$$c_{15} = 0.368630$$
$$c_{16} = -0.301462$$
$$c_{17} = 0.344614$$
$$c_{18} = -0.427782$$

If we only take account of this first zero, the price ξ_1 (in normal coordinates) would be expressed as

(2. 25)
$$\xi_1 = -0.4 z_1$$

where z_1 is some sort of (negative) standard expression for "quality", so defined that a change in this standard implies a *simultaneous* change in all the individual quality indices listed above as variates Nos. 2—7. The intensity of the change in the individual indices with a change in z_1 is defined by the coefficients (2. 24). For instance, 1 unit increase in the (negative) quality index z_1 means 0.13 units decrease in the variate No. 2 and 0.35 units increase in the variate No. 3 etc.

If we are not satisfied with this definition of the quality standard and want to study the possible *independent* influence on price exerted by some of the special quality-features, we must look for further characteristic roots. And here is where the weakness of the ellipsoid method will become apparent.

From figure 1 it looks as if there is only one other zero than B, namely one located at A. From theory we know however that there ought to be eight real zeros. The explanation is that the point A, which on the scale used in Figure 1 appears only as one zero, contains in reality seven zeros. If the vicinity around A is drawn on a larger scale we get the picture given in Figure 2. The values of the zero points are given in (2. 26).
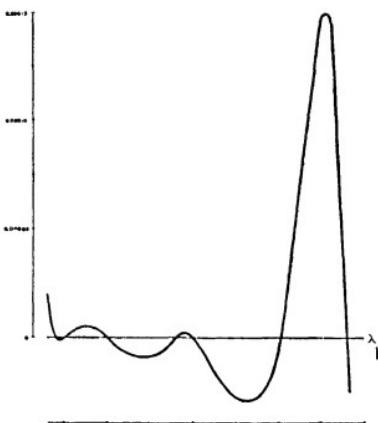
Fig. 2.

TABLE (2. 26).

CHARACTERISTIC ZEROS.

(In descending order of Magnitude.)

$\lambda = 2.67322$

1.15579

1.00443

0.7983

0.78

0.5990

0.5024

0.4851

In order to get an impression of how completely the zeros around A are mixed — from the practical view-point — we may notice that if the complete curve is drawn on the scale used in Figure 2, the bottom which in Figure 1 is marked C would be about 4 kilometres away.

The same thing may also be recognized as follows. — Roughly speaking we can say that if not more than three decimal places are significant in the values of the original correlation coefficients, anything beyond the third decimal place in the values of the characteristic polynomial is not significant. Of course, this rule is not quite exact, but is sufficient for our present purpose. Now, over the range from $\lambda = 0.48$ to $\lambda = 1.2$ where all the first seven of the characteristic roots are located, the ordinate of the characteristic polynomial $P(\lambda)$ never reaches above $+ 0.00015$ and never below $- 0.00003$. The situation is exhibited in Figure 3. Over the interval in question the ordinate of the characteristic polynomial never reaches out of the *thickness* of the heavy line in Figure 3
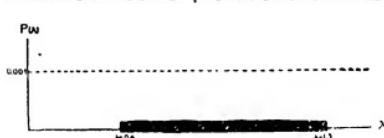
Fig. 3.

while the dotted line indicates that level which the ordinate must have reached if its value should have been significant (and even this is interpreting the significance of the original correlation coefficients rather liberally).

This means that in practice we cannot speak of certain de-

finite "zeros" of $P(\lambda)$ at all over the range from $\lambda = 0.48$ to $\lambda = 1.2$. Over this range the function $P(\lambda)$ has for all practical purposes simply a continuous contact with the λ-axis. This being the situation, the determination of the main axes of the ellipsoid — other than the one given by (2.24) — would be entirely meaningless.

This indeterminateness of the main axes may also be studied by means of certain *subcharacteristic* polynomials in the following way. The direction numbers $c_{k1}, c_{k2} \ldots c_{kn}$ for the main axes corresponding to the zero λ_k are determined by (2.17). This means that if we let

$$(2.27) \qquad\qquad G_1(\lambda) \ldots G_n(\lambda)$$

be the polynomials of λ that occur as the elements in any row of the adjoint of (2.1) (or in any column, since (2.1) is sym-metric), then the n direction numbers corresponding to a given zero λ_k is obtained simply as the values assumed by the n polynomials (2.27) when λ is put equal to λ_k.

Since (2.1) is a singular matrix in each of the zeros λ_k, and hence its adjoint of rank not higher than 1, the absolute value of the direction numbers can also be determined by putting the squares of the direction numbers proportional to the polynomials

$$(2.28) \qquad\qquad P_i(\lambda)$$

$P_i(\lambda)$ being the characteristic polynomial for the $(n-1)$ dimen-sional set obtained by leaving out the variate $no \cdot i$, $P_i(\lambda)$ may be called the subcharacteristic polynomials for the big matrix. Let

$$(2.29) \qquad\qquad \hat{r}_{ij}(\lambda)$$

be the elements of the adjoint of (2.1), considered as functions of λ. Since these functions are symmetric in (ij), and since the adjoint of a singular matrix is of rank not higher than 1, we have in any of the points $\lambda = \lambda_k$

$$(2.30) \qquad\qquad P_i(\lambda_k) \cdot P_j(\lambda_k) = [\hat{r}_{ij}(\lambda_k)]^2.$$

This shows that in any of the points $\lambda = \lambda_k$ all the subcharacter-istic polynomials defined by (2.28) must have the *same* sign.

In the point representing the smallest of the numbers λ_k, all these polynomials must even be non-negative (since for values of λ smaller than this (2. 1) is *positive* definite). The polynomials G_i need not satisfy any such sign condition since they are not *principal* (but skew) minors of (2. 1).

A high degree of indeterminateness will be attached to the main axes when the direction numbers determined either by G_i or by P_i are small. By plotting, either the curves G_i or the curves P_i over the whole range where the zeros λ_k are located, we therefore get a good impression of the degree to which the main axes have any meaning. They will be significant only to the extent that the polynomials in question (G_i or P_i) in the vicinity of the characteristic zeros reaches up above the ordinates representing one unit of the third decimal place. Figure 4 exhibits the polynomials G_i in the potato data, $G_i(\lambda)$ being here conventionally chosen as the elements in the *last* column of the adjoint of (2. 1), in other words as

(2. 31) $G_i(\lambda) = \hat{r}_{i8}(\lambda).$

The values of these polynomials in the characteristic points λ_k are given in (2: 32); and the values of the polynomials $P_i(\lambda)$ in th same points are given in (2. 23). Both these tables show very convincingly that it is only in the characteristic point $\lambda = 2.67$ that the main axes have any significance.

This confirms the conclusion that all the seven characteristic roots at A in Figure 1 must for all practical purposes be interpreted as lying in a cluster, with; out any possibility of
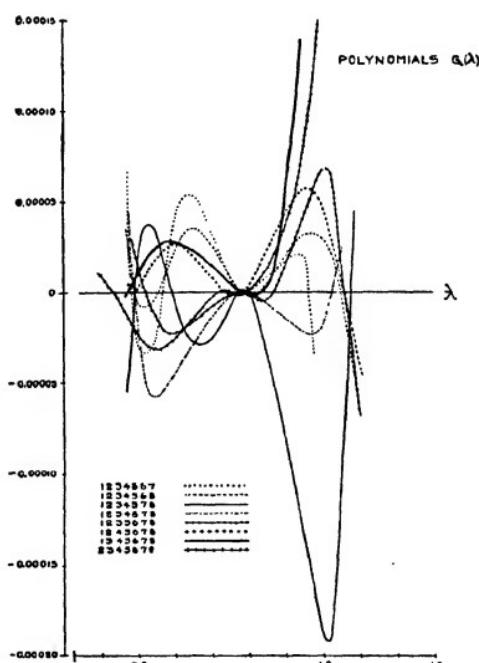
Fig. 4.

Table (2, 32). VALUES OF POLYNOMIALS $G_i(\lambda)$ IN THE CHARACTERISTIC POINTS.

λ	i = 1	2	3	4	5	6	7	8
2.673226	−15.068220	−4.896201	13.071072	−15.923427	13.952504	−11.410207	13.043494	−16.191377
1.155798	−.000095	.000348	.000014	.000528	.000725	.000603	.000299	.000275
1.004430	−.000180	.000269	−.000046	.000039	.000016	.000192	.000027	.000136
.798300	.000001	.000000	.000003	.000000	.000000	.000000	.000002	.000000
.780000	.000000	.000000	.000000	.000000	.000000	.000000	.000001	.000000
.599000	−.000025	.000028	.000028	.000022	.000050	.000006	.000024	.000039
.502400	−.000024	.000013	.000004	.000015	.000019	.000026	.000008	.000034
.485100	−.000018	.000006	.000002	.000029	.000017	.000009	.000008	.000015

Table (2, 33). VALUES OF POLYNOMIALS $P_i(\lambda)$ IN THE CHARACTERISTIC POINTS.

λ	i = 1	2	3	4	5	6	7	8
2.673226	−14.023037	−1.480574	−10.552085	−15.659916	−12.023201	−8.040853	−10.507592	−16.191299
1.155798	.000032	.002602	.000006	.001013	.001908	.001320	.000324	.000274
1.004430	−.000236	.000531	.000014	.000035	.000002	.000267	.000004	.000137
0.798300	.000001	.000001	.000027	.000001	.000000	.000001	.000012	.000000
0.780000	−.000006	.000001	.000008	.000001	.000000	.000011	.000019	.000000
0.599000	−.000016	.000020	.000019	.000013	.000365	.000002	.000015	.000039
0.502400	−.000019	.000005	.000000	.000005	.000011	.000020	.000002	.000034
0.485100	.000027	.000003	.000001	.000064	.000019	.000005	.000004	.000015

discriminating between them. Since it is out of the question to assume them all to be significantly different from zero, the conclusion of the ellipsoid analysis must be that it is only the root at B that can be looked upon as significantly different from zero. Consequently the price must by this method be looked upon as capable of being expressed in terms of *one single quality index* as indicated in (2. 25).

This result is however in contradiction both with the concrete knowledge we have of the data and with the impression one gets by a detailed *cross classification* on the individual observations. The general knowledge of the data as well as the cross classification indicates that there is more than one significant degree of freedom. Some further support for this conclusion will also be found by the analysis of Section 31 based on the "bunch" technique. The bunch analysis will also give some indication of which one of the observational variates are the most important to include.

In view of this negative conclusion regarding the usefulness of the ellipsoid method I decided to postpone publication of my work on the ellipsoid method until I had a better method to offer by which the ellipsoid method could be compared[1].

3. MINIMAL λ-ROOTS AND MAXIMAL ζ-ROOTS.

If the characteristic roots are to be used at all as criteria of confluency, it seems better to compute only one of them, namely the *minimal* root, but then do this for all possible subsets. By doing this one dodges at least that difficulty which consists in judging whether the roots are significantly different from one another. The computation of the minimal roots gives an alternative set of parameters to study instead of the scatterances, each scatterance is simply replaced by its minimal root. For the systematic computation of these roots in all possible subsets I have worked out the following technique which has been found convenient and has been used in all the numerical work done along this line at the Oslo Institute.

[1] On other types of data the ellipsoid method may perhaps be used with advantage. Professor Harold Hotelling in his highly interesting paper »Analysis of a Complex of Statistical Variables into Principal Components», The Journal of Educational Psychology, 1933, has applied it to psychological data. In this paper he also gives a method of successive approximation to the characteristic roots which will work well, it seems, whenever the roots are significantly distinct. I wonder whether his method will work equally well when the roots lie close together as in the above eight variate problem.

Instead of considering the minimal λ-roots, it is for the practical computations more convenient to consider the corresponding maximal ζ-roots. We shall therefore consider the polynomial $Q(\zeta)$ defined by (2.3) instead of the polynomial $P(\lambda)$ defined by (2.1).

Let ζ and ζ_i be the maximal roots of $Q(\zeta)$ and $Q_i(\zeta)$ respectively, $Q_i(\zeta)$ being the subcharacteristical polynomial obtained by leaving out the variate No $\cdot i$.

The shape of the functions $Q(\zeta)$ and $Q_i(\zeta)$ in the first range to the left of unity must be as indicated by the curves A and B respectively in Figure 5. (The curve A in Figure 5 is constructed by the formula $Q(\zeta) = 0.25 + 0.37\,\zeta - 1.53\,\zeta^2 + \zeta^4$, the curve B by the formula $Q_i(\zeta) = -0.15 - 0.59\,\zeta + \zeta^3$. The numerical constants here refer to an actual example). We are in particular interested in the relative location of the zeros.

In order to study this we first notice that (2.1) is a positive definite determinant for $\zeta = 1$ and all the minors continuous

Fig. 5.

in ζ, therefore as ζ decreases from 1, $Q(\zeta)$, when written as determinant, must, to begin with, remain positive definite. As we move from $\zeta = 1$ towards the left, none of the polynomials Q_i can therefore vanish before Q. In other words: *The maximal roots in a set of variates is never less than the maximal root in any of its subsets.*

As we move towards the left, there will come a point where for the first time one of the polynomials Q_i vanish. Let $\zeta = \zeta^0$ be the point where this happens. Differentiating the determinant (2.1) we get

$$(3.1) \qquad Q'(\zeta) = \frac{dQ(\zeta)}{d\zeta} = Q_1(\zeta) + \ldots + Q_n(\zeta).$$

This shows that, up to the point ζ^0, Q must certainly be monotonically decreasing with decreasing ζ. Consequently: If ζ is the maximal root in a certain set, and ζ^0 the maximal root in that one of its subsets which has the largest maximal root, we have
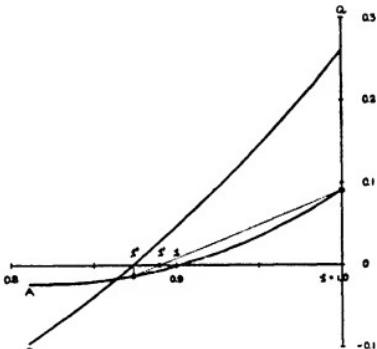
(3.2) $$\zeta^0 \gtreqless \zeta \gtreqless 1.$$

Furthermore: *In the range (3.2) the polynomial $Q(\zeta)$ has exactly one zero.* This gives a very convenient method of successively approximating to the maximal roots in all possible subsets of a given big set. First the maximal roots in the three dimensional sets are computed directly by the usual formula for the solution of a third degree equation. This is easy since the second power of ζ is lacking in the equation. For higher sets one then proceeds as follows. Since Q is non-negative in the point $\zeta = 1$ and non-positive in $\zeta = \zeta^0$ and has exactly one zero in this range, we can by linear interpolation take

(3.3) $$\zeta' = \zeta^0 - \frac{\zeta^0 - 1}{Q(\zeta^0) - Q(1)} \, Q(\zeta^0)$$

as a first approximation to the minimal root in the bigger set. Of course $Q(1)$ is nothing but $\triangle$. A second approximation will be

(3.4) $$\zeta'' = \zeta' - \frac{\zeta' - \zeta^0}{Q(\zeta') - Q(\zeta^0)} \, Q(\zeta')$$

etc.

The computations are ordered in a difference scheme as indicated in (3.5).

Table (3.5). DIFFERENCE SCHEME FOR THE COMPUTATION OF MAXIMAL ROOTS.

ζ	Q	Difference of ζ	Difference of Q	Divided difference of ζ with respect to Q
1.000	$\triangle$			$\dfrac{\zeta^0 - 1.000}{Q(\zeta^0) - \triangle}$
ζ^0	$Q(\zeta^0)$	$\zeta^0 - 1.000$	$Q(\zeta^0) - \triangle$	
ζ'	$Q(\zeta')$	$\zeta' - \zeta^0$	$Q(\zeta') - Q(\zeta^0)$	etc.
ζ''	$Q(\zeta'')$	$\zeta'' - \zeta'$	$Q(\zeta'') - Q(\zeta')$	

The computation of the values needed of the polynomial $Q(\zeta)$ in Table (3.5) are done most conveniently by the formula (2.10).

The scatterances and minimal roots have been computed at the Institute for a large series of data, primarily American con-

sumption statistics, which were collected and brought in shape for this analysis by Dr. Waugh.

If the scatterances and minimal roots shall have any use as tests of linear confluency, the conclusions reached by using these two different sets of criteria ought to be essentially the same. Although no exact correspondence was found, yet in most cases, the discrepancy was not very great. In general there was a definite *tendency* for the maximal ζ root to be large when the scatterance was small and vice versa. The kind of information yielded by these two sets of parameters may therefore be looked upon as roughly the same.

The graph in Figure 6 indicate the connection between scatterances and maximal roots in a set of the above mentioned American consumption data. (Butter 1919—31).
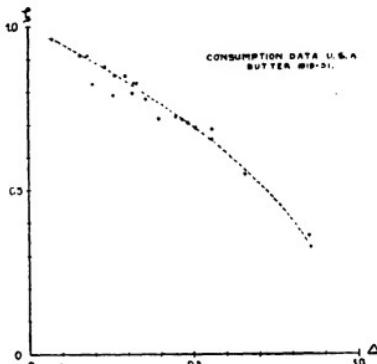
Fig. 6.

4. LEMMAS ON CERTAIN PROPERTIES OF DETERMINANTS. THE HEAD COEFFICIENTS.

In the following analysis we shall have to make use of a few properties of determinants which it will be convenient to indicate here[1].

(2. 1) with the expansion (2. 2) is the characteristic polynomial for the matrix r_{ij}. The adjoint of r_i, namely the matrix

(4. 1) $$\hat{r}_i$$

also has a characteristic polynomial. Let it be $\hat{P}(\lambda)$, and let $\hat{A}$ be its coefficients. Then it is a classical fact in matrix algebra (contained as a special case in Frobenius theorem) that

---

[1] A more complete summary of the classical facts in quadratic forms and matrix algebra that are of particular interest from the wiew point of application to economics and statistics, were given in my Colloquium Lectures at the meeting of the Econometric Society in Leyden 1933. A mimeographed account of these lectures based on my own notes and of notes taken by Mr. M. H. Belz during the lectures, is available and may be ordered through the Economic Institute, Oslo. (Price Kr. 4.00.)

(4. 2)
$$\hat{P}(\lambda) = (-)^n \frac{\lambda^n}{A_n} P\left(\frac{A_n}{\lambda}\right).$$

In other words we have

(4. 3)
$$\hat{A}_{n-k} = A_k A_n^{n-k-1}.$$

As an example of (4. 2) we may take the case $n = 2$. Here the characteristic polynomial is

$$P(\lambda) = \begin{vmatrix} r_{11} - \lambda & r_{12} \\ r_{21} & r_{22} - \lambda \end{vmatrix} = \triangle_{12} - (r_{11} + r_{22})\lambda + \lambda^2.$$

And the adjoint characteristic polynomial is

$$\hat{P}(\lambda) = \begin{vmatrix} r_{22} - \lambda & -r_{12} \\ -r_{21} & r_{11} - \lambda \end{vmatrix} = \triangle_{12} - (r_{11} + r_{22})\lambda + \lambda^2.$$

In other words for $n = 2$, the characteristic polynomial and the adjoint characteristic polynomial coincide, which cheks with (4. 2).

In the case $n = 3$ the adjoint characteristic polynomial is

$$\hat{P}(\lambda) = \begin{vmatrix} [(r_{22}r_{33} - r_{32}r_{23}) - \lambda] & -(r_{12}r_{33} - r_{32}r_{13}) & (r_{12}r_{23} - r_{22}r_{13}) \\ -(r_{21}r_{33} - r_{31}r_{23}) & [(r_{11}r_{33} - r_{31}r_{13}) - \lambda] & -(r_{11}r_{23} - r_{21}r_{13}) \\ (r_{21}r_{32} - r_{31}r_{22}) & -(r_{11}r_{32} - r_{31}r_{12}) & [(r_{11}r_{22} - r_{21}r_{12}) - \lambda] \end{vmatrix}$$

Writing this determinant out and collecting the terms we find that the expression reduces to

$$\hat{P}(\lambda) = \triangle^2_{123} - \triangle(r_{11} + r_{22} + r_{33})\lambda + (\triangle_{12} + \triangle_{13} + \triangle_{23})\lambda^2 - \lambda^3$$

which also checks with (4. 2).

For various purposes the three first coefficients of the characteristic polynomial are of special interest; we shall call them *the head coefficients* and designate them by special letters

(4. 4)
$$\begin{aligned} A_n &= \triangle \\ A_{n-1} &= \varPhi \\ A_{n-2} &= \varPsi \end{aligned}$$

Thus $\triangle$ is the determinant (1. 15) itself, $\varPhi$ the sum of all its

$n\,(n-1)$ rowed principal minors, and Ψ the sum of all its $\binom{n}{2}\,(n-2)$ rowed principal minors. In a given subset $(\alpha,\beta,\ldots\gamma)$ we may use the notation $\Delta_{\alpha\,\beta\,\ldots\gamma}$ etc. And in the set obtained from a given set by leaving out, say the subscript, α, we may use the notation $\Delta_{)\alpha(}$ etc.

This being so, let us consider any set of ν affixes $(\alpha,\beta\ldots\gamma)$ and let us form the adjoint correlation matrix $\hat{r}_{ij(\alpha,\,\beta\,\ldots\gamma)}$, the adjunction being taken within the set $(\alpha,\beta\ldots\gamma)$. For shortness let us denote this ν-rowed matrix

$$(4.5)\qquad\qquad s_{ij}=\hat{r}_{ij(\alpha,\,\beta\,\ldots\gamma)}$$

Consider the characteristic polynomial for the ν-rowed matrix s, let $\triangle^{(s)}$ and $\varPhi^{(s)}$ be its first two head coefficients. Further, let $\triangle^{(s)}_{)p(}$ and $\varPhi^{(s)}_{)p(}$ be the head coefficients in the $\nu-1$ rowed matrix obtained from s by leaving out the row No. p and the column No. p. We shall study the difference $\varPhi^{(s)}\,\triangle^{(s)}_{)p(}-\triangle^{(s)}\,\varPhi^{(s)}_{)p(}$.

To evaluate this expression we need Sylvester's formula on the minors in an adjoint. This formula may be written

$$(4.6)\qquad \left|\hat{a}\right|_{\underbrace{h\,k\,\ldots\,l}_{m}\,\cdot\,\underbrace{u\,v\,\ldots\,w}_{m}}=\varepsilon\cdot\left|a\right|_{)u\,v\,\ldots\,w\,\cdot\,h\,k\,\ldots\,l(}\cdot\left|a\right|^{m-1}$$

ε being the signfactor $(-1)^{h+k+\cdots+l+u+v+\cdots+w}$, $\left|a\right|$ is a given n-rowed determinant, the expression in the left member of (4.6) stands for the m-rowed determinant consisting of the rows Nos. $h,k\ldots l$ and the columns Nos. $u,v\ldots w$ from the adjoint of a. The first factor in the right member indicates the $n-m$ rowed determinant which is obtained when the rows $u,v\ldots w$ and the columns $h,k\ldots l$ are omitted from $\left|a\right|$ itself; the second factor in the right member is simply the $(m-1)th$ power of the n-rowed determinant $\left|a\right|$.

This being so consider $\triangle^{(s)}_{)p(}$. It may be looked upon as obtained in the following way. We take as a starting point the ν-rowed matrix of the original correlation coefficients r_{ij} in the set $(\alpha,\beta\ldots\gamma)$. Of this matrix we take the adjoint, the adjunction being made within the set $(\alpha,\beta\ldots\gamma)$. In this ν-rowed adjoint we consider the $(\nu-1)$-rowed minor obtained by leaving out the row and the column p.

By Sylvester's formula this minor is equal to $r_{pp}\,\Delta^{\nu-2}$. Further,

by (4. 3) we have $\Phi^{(s)} = \triangle^{\nu-2} \sum_\varkappa r_{\varkappa\varkappa}$ ($\varkappa$ running through $\alpha, \beta \ldots \gamma$).
On the other hand we have $\Phi^{(s)}_{)p(} = \sum_\varkappa |s|_{)p\varkappa \cdot p\varkappa(}$, which by Sylvester's
formula is equal to $\sum_\varkappa \begin{vmatrix} r_{pp} & r_{p\varkappa} \\ r_{\varkappa p} & r_{\varkappa\varkappa} \end{vmatrix} \cdot \triangle^{\nu-3}$. Finally by Sylvesters's
formula $\triangle^{(s)} = \triangle^{\nu-1}$, so that

$$(4.\ 7) \quad \begin{aligned} \Phi^{(s)} \triangle^{(s)}_{)p(} - \triangle^{(s)} \Phi^{(s)}_{)p(} &= \triangle^{2\nu-4}[r_{pp}\sum_\varkappa r_{\varkappa\varkappa} - \sum_\varkappa(r_{pp}r_{\varkappa\varkappa} - r^2_{p\varkappa})] = \\ &\triangle^{2\nu-4}\sum_\varkappa r^2_{p\varkappa}. \end{aligned}$$

Now let us take the adjoint of the ν-rowed matrix s_{ij}. Apart from a factor of proportionality this brings us back again to the original matrix r_{ij}; more precisely we have

$$(4.\ 8) \qquad\qquad \hat{s}_{ij} = r_{ij} \cdot \triangle^{\nu-2}.$$

Inserting this in the right member of (4. 7) this member takes on the form $\sum_\varkappa \hat{s}^2_{p\varkappa}$. The formula thus obtained holds good for any matrix s, it consequently also holds good for the original matrix r, hence

$$(4.\ 9) \qquad\qquad \sum_\varkappa \hat{r}^2_{p\varkappa} = \Phi \triangle_{)p(} - \triangle \Phi_{)p(}$$

where the head coefficients $\triangle$ and Φ are now taken in the set $\alpha, \beta \ldots \gamma$ of the original correlation matrix; $\varkappa$ in (4. 9) runs through $\alpha, \beta \ldots \gamma$.

As an example consider a three-rowed correlation matrix. The elements in the first row of the adjoint are here

$$\hat{r}_{11} = 1 - r^2_{23}, \quad \hat{r}_{12} = -(r_{12} - r_{13}r_{23}), \quad \hat{r}_{13} = (r_{12}r_{23} - r_{13})$$

The sum of the square of these three quantities is

$$1 - 2r^2_{23} + r^4_{23} + r^2_{12} - 2r_{12}r_{13}r_{23} + r^2_{13}r^2_{23} + r^2_{12}r^2_{23} - 2r_{12}r_{13}r_{23} + r^2_{13}$$

On the other hand the right member of (4. 9) is in this case

$$(1 - r^2_{12} + 1 - r^2_{13} + 1 - r^2_{23}) \cdot (1 - r^2_{23}) - 2[1 + 2r_{12}r_{13}r_{23} - (r^2_{12} + r^2_{13} + r^2_{23})].$$

It is easily seen that these two expressions are equal.

By taking the sum of (4. 9) over p and noticing that

$$(4.\ 10) \qquad\qquad \begin{cases} \sum_p \triangle_{)p(} = \Phi \\ \sum_p \Phi_{)p(} = 2\Psi \end{cases}$$

we get

$$(4.11) \qquad \sum_{ij} \hat{r}_{ij}^2 = \varPhi^2 - 2\,\varPsi\,\triangle$$

In the summation in (4.11) i and j run independently of each other through all the subscripts $(\alpha, \beta \ldots \gamma)$ in the set in which the adjunction $\hat{r}$ is taken; and the head coefficients in the right member refer to the same set $(\alpha, \beta \ldots \gamma)$. (4.11) may be used as a convenient check on the computation of the sumsquares in the individual rows computed according to (4.9).

5. REGRESSION SPREADS.

The minimising of the sumsquares of the deviation from the regression plane may be done in different directions, for instance in the direction of the x_1 axis, in that of the x_2 axis, etc. These regressions we shall call the elementary regressions. The coefficients of these regressions (when the variates are taken in the normalised form) are nothing but the elements $\hat{r}_{ij}$ of the adjoint correlation matrix. The elements in the first row of this matrix are the coefficients of the first elementary regression, those of the second row the coefficients of the second elementary regression, etc.

The fact that these various sets of coefficients are nearly proportional (in other words that all the elementary regression planes nearly coincide) one would intuitively take as a sign that the regression plane determined in this set of variates is significant. I shall later discuss this idea more closely; for the moment let us adopt it heuristically.

A first condition that would have to be fulfilled in order that these n sets of coefficients are proportional is obviously that these sets, when considered as defining n points in an n-dimensional scatter diagram, lie in a plane through origin of this space. The idea therefore suggests itself to measure the degree of conformity between the n elementary regression planes by computing the scatterance of the regression coefficients considered as statistical observations. These "observations" should however not be reduced to their mean values before computing the scatterance, since the "regression" plane in these "observations" *ought* to go through origin.

This leads to computing the expression

$$(5.1) \qquad \frac{\begin{vmatrix} \hat{R}_{11} \ldots \hat{R}_{1n} \\ \cdots\cdots\cdots\cdots \\ \hat{R}_{n1} \ldots \hat{R}_{nn} \end{vmatrix}}{\hat{R}_{11}\,\hat{R}_{22}\ldots\hat{R}_{nn}}$$

where

$$(5.2) \qquad \hat{R}_{ij} = \sum_k \hat{r}_{ik}\hat{r}_{kj}.$$

The accent $\frown$ on the magnitudes R in the above formula may be interpreted as true adjunction symbols. Indeed, by virtue of the formula for symbolic matrix multiplication we may define (R_{ij}) as the *square* of the matrix r_{ij}, in other words

$$(5.3) \qquad R_{ij} = \sum_k r_{ik}r_{kj},$$

$\hat{R}_{ij}$ is then the adjoint of R_{ij}.

Similarly we may compute the corresponding coefficient (5.1) for any subset $(\alpha, \beta \ldots \gamma)$ in the big set $12 \ldots n$. In this latter case the adjunction sign $\frown$ will of course have to be interpreted as adjunction within the subset $(\alpha, \beta \ldots \gamma)$.

By the formulae of the preceding Section the expression (5.1) may be considerably reduced. As a matter of fact it may be expressed very simply by the head-coefficients of the character-istic polynomial. We first notice that the numerator of (5.1), namely the determinant $|\hat{R}|$ by Sylvester's formula is $|R|^{n-1}$, but the matrix R was the symbolic square of r, so that

$$(5.4) \qquad |\hat{R}| = \Delta^{2(n-1)}.$$

Further, each factor in the denominator of (5.1) is by (4.9) equal to

$$(5.5) \qquad \hat{R}_{pp} = \Phi \cdot \Delta_{)p(} - \Delta \cdot \Phi_{)p(}$$

The expression (5.4) shows that, in order to get a coefficient comparable to the scatterance, we ought to take the $2(n-1)th$ root of (5.1). We thus finally get

$$(5.6) \qquad \frac{\Delta}{\sqrt[2(n-1)]{\hat{R}_{11}\hat{R}_{22}\ldots\hat{R}_{nn}}}$$

where the $\hat{R}_{pp}$ are given by (5.5).

The coefficient defined by (5.6) may be called *the regression spread*. It may be looked upon as a sort of corrected scatter-ance, the denominator of (5.6) being the correction factor.

An important property of the coefficient (5.6) is that it is capable of *increasing* if more variates are included. In this respect it differs from the scatterance.

If the regression spread increases as a new variate is included, it may be taken as a warning signal that the inclusion of the new variate is not warranted. This criterion is not, however, final. The behavior of the regression spreads must be scrutinized by principles similar to those used in the study of the scatterances (see Section 1).

6. LINE COEFFICIENTS.

The regression spread defined in Section 5 gives an expression for the closeness with which the adjoint of a given correlation matrix comes to being *singular*, but if we are looking for a unique regression plane it would in fact be more plausible to construct a criterion for the adjoint correlation matrix being of rank not higher than 1 (i. e. all the rows proportional).

If the adjoint correlation matrix — so far as the systematic variations are concerned — is of rank 1, all the two rowed determinants $\hat{r}_{ii}\,\hat{r}_{jj} - \hat{r}_{ij}^2$ ought to be small. Therefore, if the average values of these determinants taken over all the subsets (ij) are small as compared with the average value of their leading terms $\hat{r}_{ii}\,\hat{r}_{jj}$, we may take it as an indication that the matrix $(\hat{r}_{ij})$ is close to being of rank not higher than 1. The averages here considered only consist of non-negative elements (since (r_{ij}) is positive definite), so that there is no danger of the average becoming small because positive and negative terms cancel out. The sum of the two-rowed determinants considered is nothing but $\hat{A}_2 = A_{n-2}\,A_n = \Psi \Delta$. And the sum of the leading terms is

$$\sum_{i<j} \hat{r}_{ii}\,\hat{r}_{jj} = \tfrac{1}{2}\,[\sum_i \sum_j \hat{r}_{ii}\,\hat{r}_{jj} - \sum_{\mathsf{x}} \hat{r}_{\mathsf{x}\mathsf{x}}^2]$$

which reduces to

$$\tfrac{1}{2}\,[(\sum_i \triangle)_{i(})^2 - \sum_i \triangle_{)i(}^2]$$

Hence the smallness of the expression

(6 1)
$$\frac{2\,\Psi\,\Delta}{\Phi^2 - \sum_i \triangle_{)i(}^2}$$

may be taken as a criterion for the adjoint correlation matrix being of rank 1. The expression (6. 1) may be called the *line coefficient*, because it expressed the closeness with which the normals of the various elementary regression planes come to lying in a common line. The components of these normals are of course nothing but the elements in the various rows of the adjoint correlation matrix.

By (4. 11) the expression for the line coefficient may also be written

$$(6. 2) \qquad \frac{\Phi^2 - \sum_{ij} \hat{r}_{ij}^2}{\Phi^2 - \sum_i \hat{r}_{ii}^2}.$$

Since Ψ and $\triangle$ are non-negative and the denominator of (6. 1) was originally written as a double sum of the produkt $\hat{r}_{ii} \cdot \hat{r}_{jj}$, which is non-negative, it is seen that (6.1) is larger than 0. On the other hand (6.2) shows that it must be not larger than unity since the term that is subtracted in the numerator is not less than the one that is subtracted in the denominator. In other words, the coefficient considered must lie between zero and unity.

If it is wanted to compute the line coefficients in all possible subsets, the work can most conveniently be arranged as follows. The $\triangle$ in all subsets are supposed to be computed (for instance by the tilling technique of Section 15). The second head coefficient Φ is then computed in all subsets by the recurrence formula

$$(6. 3) \qquad \Phi_{\alpha\beta\ldots\gamma} = \sum_{i = \alpha\beta\ldots\gamma} \triangle_{\alpha\beta\ldots)i(\ldots\gamma}$$

$a, \beta \ldots \gamma$ being a given ν-set. This formula is only a special case of (2. 4). For each ν-level the sum of all the Φ thus computed, are checked by

$$(6. 4) \qquad \begin{aligned} \sum_{\alpha < \beta < \ldots < \gamma} \Phi_{\alpha\beta\ldots\gamma} &= (n - \nu + 1) A_{\nu-1(12\ldots n)} = \\ &= (n - \nu + 1) \sum_{h=0}^{\nu-1} \binom{n-h}{\nu-h-1} B_{h(12\ldots n)}. \end{aligned}$$

To apply this check, either the A's or the B's in the total set must be computed directly; this is a simple matter when the $\triangle$'s are already listed in a systematic way, for instance if they

are computed by the tilling technique as decribed in Section 15, they can then quickly be copied on a listing adding machine directly from the tilling tables, (6.4) is only a special case of (2.5), the last member in (6.4) is obtained by using (2.7).

The recurrence computations (6.3) are most easily arranged by keeping the $\triangle$'s written in tables with uniform spacing and in the exact concentric order defined in Section 15. If a listing adding machine is used for the determination of the A's in the big set, such a lists of the $\triangle$ are already available. On these lists no indication of *which* scatterance the various figures refer to should be given, but one should rely on the uniform spacing and the exact order defined by the concentric numbering. In order to pick out the correct figures to go into the recurrent computation (6.3) we use "combination strips", that is strips of paper or cardboard with check marks in positions that indicate the figures to be used for any given subset $(\alpha, \beta \ldots \gamma)$. A set of such combination strips is very useful also for other recurrence computation of a similar kind. (See for instance (6.5) etc. below). We have found it both safer and quicker to rely on the uniform spacing and the combination strips than to write down on the lists of the $\triangle$'s the subset to which each $\triangle$ belong.

When the Φ in all subsets are computed and checked, the quantities 2Ψ are computed recurrently by

$$(6.5) \qquad 2\psi_{\alpha\beta\ldots\gamma} = \sum_{i=\alpha\beta\ldots\gamma} \Phi_{\alpha\beta\ldots)i(\ldots\gamma}$$

and checked by

$$(6.6) \qquad \sum_{\alpha<\beta\ldots<\gamma} 2\Psi_{\alpha\beta\ldots\gamma} = 2\binom{n-\nu+2}{2} A_{\nu-2(12\ldots n)} =$$
$$= 2\binom{n-\nu+2}{2} \sum_{h=0}^{\nu-2} \binom{n-h}{\nu-h-2} B_{h(12\ldots n)}.$$

This check is obtained by putting $k = \nu - 2$ in (2.5). The check on ψ is, of course, done for each ν-level in the same way as for Φ.

Further the squares $\triangle^2$ and Φ^2 are computed in all subsets, and the sumsquares

$$(6.7) \qquad D_{\alpha\beta\ldots\gamma} = \sum_{i=\alpha\beta\ldots\gamma} \triangle^2_{\alpha\beta\ldots)i(\ldots\gamma}$$

are computed and checked in each subset. This is done in precisely the same way as the Φ were built up from the $\triangle$

[compare (6.3) with (6.7)]. The checking formula written out will now be

$$(6.8) \qquad \sum_{\alpha < \beta \ldots < \gamma} D_{\alpha\beta \ldots \gamma} = (n - \nu + 1) \sum \triangle^2_{(\nu - 1)\text{-rowed}}.$$

The sum in the left member of (6.8) denotes the sum of all the D's computed on the level ν by means of (6.7), and the sum in the right member is the sum of all the $\binom{n}{\nu-1}$ squares of the $(\nu-1)$-rowed $\triangle$'s contained in the total set. This latter sum must be computed directly in order to apply the check. All the work (6.5) and (6.7) is done by the combination strips.

By means of the above quantities the line coefficients in all sets are easily computed by (6.1). The difference $\omega^2 - D$ which is in the denominator of (6.1), may of course be verified by a simple checksum utilizing the sum of ω^2 and the sum of D.

The practical use of the line coefficients is similar to that of the regression spread and the scatterances.

In a sense the use of the line coefficients and the regression spreads is a little safer than that of the scatterances because they will actually increase when we get into multiple collinear sets, while the scatterances will always decrease. On the other hand the situation may be such that we must accept a set of variates even though it shows some increase in the regression spread or line coefficient. Whether or not a slight increase in any of these parameters should be accepted cannot be definitely decided unless by the more elaborate method of Part III. We shall later discuss numerical examples which will illustrate both the advantages and the limitations of the line coefficients.

PART II. SCATTER FUNCTIONS AND »TRUE» REGRESSIONS.

7. INSIDE AND OUTSIDE INFLUENCES. SYSTEMATIC COMPONENTS AND DISTURBANCES. THE NOTION OF DISTURBING INTENSITY.

Although the various empirical test-parameters considered, scatterances, characteristic roots, beam coefficients, etc. throw *some* light on the question of confluency, they do not in all cases furnish a conclusive criterion.

To advance any further in the matter it seems that we need a more systematic analysis based on certain definite assump-

tions about the *nature* of the variates. The present Part II is concerned with this type of analysis. It leads up to a certain statistical technique: the bunch analysis which will be discussed in the two following parts.

Let us assume that each of the *observed* variates $x_1 \ldots x_n$, depends on certain other "basic" variates $y_1 \ldots y_M$ which are not observed, but which have actually been present and have determined each of the observed results. If the connection is linear, the dependency can be written in the form

$$(7.1) \qquad x_i = \textstyle\sum_K p_{iK} y_K$$

where the p_{iK} are constants. The n rowed and M columned matrix

$$(7.2) \qquad \| p_{iK} \| = \left\| \begin{matrix} p_{11} \cdots p_{1M} \\ \cdots\cdots\cdots \\ p_{n1} \cdots p_{nM} \end{matrix} \right\|$$

characterises the way in which the observational variates are built up from the basic variates.

For the moment we do not specify any further the concrete nature of the causal relations between the y's and the x's, we just take (7.1) as a conceptual pattern which may serve as a starting point for the analysis.

In many cases it seems plausible to formulate the conception of the basic variates in such a way that they become *uncorrelated*. In this case any amount of correlation that has been observed between the x's will be due simply to the fact that one or more of the y's occur in more than one x. However, in practice — for instance if one actually tries to produce a numerical example illustrating a connection of the form (7.1) — it is not plausible to assume *exact* non-correlation between the basic variates. In many cases it will be necessary to work with the more or less vague assumption that the correlation between these variates shall be "small", or that it shall be determined only by "random" fluctuations.

We assume the x's as well as the y's to be measured from their means. Furthermore, it does not restrict generally if we assume all the basic variates to have unit sumsquare. This only involves a corresponding interpretation of the coefficients p_{iK}. In other words we assume that

(7.3) $$[y_K^2] = 1 \qquad (K = 1, 2 \ldots n)$$

where [ ] denotes a summation over all the observations. When the assumption (7.3) is made, the coefficients p_{iK} express the relative importance of the various basic variates in the determination of a given observed variate.

If (7.3) is fulfilled, the cross moment of the variates y_H and y_K is the same as the correlation coefficient between them, namely

(7.4) $$s_{HK} = \frac{[y_H y_K]}{\sqrt{[y_H^2][y_K^2]}}.$$

The crossmoment between the observed variates x_i and x_j will be

(7.5 $$m_{ij} = [x_i x_j] = \sum_{HK} s_{HK} p_{iH} p_{jK}$$

H and K running independently of each other through all the numbers $1, 2 \ldots M$. In other words the cross moment m_{ij} is nothing but the value assumed by the bilinear form built over the matrix s_{HK}, when the variates in this form are put equal to the $i - th$ and j - th row in (7.2) respectively.

If the basic variates are assumed exactly noncorrelated, (7.5) reduces to

(7.6) $$m_{ij} = \sum_K p_{iK} p_{jK}.$$

If we further assume that the units of measurement are chosen so that the observational variates satisfy a relation analogous to (7.3), (7.6) will at the same time be the correlation coefficient between the observational variates x_i and x_j. If we finally assume that $M < n$, we get the special hypothesis regarding the correlation coefficients on which Spearman's two factor theory and Thurstones multifactor theory are built. Here we shall not follow up this line of approach.

In general we shall assume that M may be a number larger than n. A great number of different cases must then be envisaged according to *how* an interconnection between the x's may be produced through the y's.

Consider first two observational variates x_i and x_j. The basic variates y may then be classified in three groups accord-

ing as they occur in both variates x_i and x_j, in only one of them, or in none of them. Those basic variates that occur in both observational variates x_i and x (with a coefficient p different from zero) will be said to constitute an *inside* or *systematic* influence on the observed interconnection between x_i and x_j, while those basic variates that occur in one of the two observational variate, but not in the other, will be said to constitute an *outside* influence, or an accidental disturbance on the variation of the observational variate in question. Those basic variates that do not occur in either of the two observational variates are of no interest in this connection.

If there is at least one basic variate that thus occurs in both the observational variates considered, we shall say that there exists an inside or systematic connection between these two observational variates, otherwise they will be said to be systematically unconnected. The connection here considered may be called inside or systematic *in the restricted sense* to distinguish from another sort of inside influence to be considered presently.

Next, consider a set of ν observational variates, say

(7. 7)   Nos. $\alpha, \beta \ldots \gamma$.

A given basic variate may now be classified in one of three groups according as it occurs in more than one of the observational variates in this set, or in just one, or in none. In the first case the variate in question will be said to exert an influence that is *inside* the set $(\alpha, \beta \ldots \gamma)$ or is *systematic* with respect to this set; in the second case the basic variate will as before be said to represent a disturbance. The third case has no interest in connection with the set $(\alpha, \beta \ldots \gamma)$.

Thus, as we enlarge the set of observational variates, the larger will be the portion of their determining factors that must be considered as "inside" or "systematic", and the smaller will be the part that is still left in the category of "disturbances". Furthermore, if we include in the statistical analysis some new observational variate in order to explain some of the lack of fit which we had in the original set, we must not forget that in practice this always means that we introduce a new *complex*, of which perhaps only a small part is systematically connected with those observational variates we originally considered. It

is just this fact that creates the real problems of confluence analysis.

There are many reasons which prevent us from ever reaching a situation where all possible fluctuations are exactly taken account of by those variates we have included in the analysis. A smaller or larger part of the fluctuations must always be left in the indiscriminated category of "disturbances". In each particular problem nature has drawn a more or less distinct line of demarcation between those fluctuations which we may hope to explain by studying simultaneously certain specified observational variates and the inexplicable rest. This justifies speaking of "systematic variation" and "random disturbances" in a semi-absolute sense, the set of variates within which the expressions shall be understood being defined in a more or less precise way by the very nature of the problem considered. As a rule when we use the terms "systematic variates" and "disturbances" without further specification, we take them in this semi-absolute sense.

If in some way or another a distinction has been drawn between those basic variates which create the inside and those that create the outside influence, we may write each observational variate x_i as the sum of two components

$$(7.8) \qquad x_i = x_i' + x_i''$$

x_i' being the systematic part of x_i — that is the part that is systematically connected with the other variates considered — and x_i'' its "disturbance".

Let $t = 1, 2 \ldots N$ be numbering of the observations, and let, as before, the Gaussian symbol [ ] denote a summation over all the observations.. Since we assume all the variates to be measured from their means, we have $[x_i] = [x_i'] = [x_i''] = 0$.

The observed cross moment $m_{ij} = [x_i x_j]$ is equal to

$$(7.9) \qquad m_{ij} = m_{ij}' + [x_i' x_j''] + [x_i'' x_j'] + [x_i'' x_j'']$$

where m_{ij}' is the cross moment of the systematic parts of the variates i and j.

For $i = j$ (7.9) reduces to

$$(7.10) \qquad m_{ii} = m_{ii}' + m_{ii}'' + 2[x_i' x_i'']$$

The decomposition (7.8) and the ensuing formulae (7.9) and (7.10) have here been derived from the notion of basic variates and the formula (7.1). But there is obviously nothing that prevents us from taking the decomposition (7.8) independently as the starting point of the analysis.

There are three assumptions which may suggest themselves regarding the nature of the correlation between the variates involved in (7.9) and (7.10).

I. There is no correlation between the accidental parts of two different observational variates.

II. There is no correlation between the accidental part of one observational variate and the systematic part of another.

III. There is no correlation between the accidental and systematic parts of a given observational variate.

If the assumptions (I) and (II) are made we get

$$(7.11) \qquad m_{ij} = m'_{ij} \text{ for } i \neq j$$

and if the assumption (III) is made we get

$$(7.12) \qquad (1 - \lambda_i) m_{ii} = m'_{ii}$$

where

$$(7.13) \qquad \lambda_i = \frac{m''_{ii}}{m_{ii}}$$

λ_i may be called the *disturbing* intensity of the variate x_i; it is the ratio between the sumsquare of the disturbance of x_i and the sumsquare of the observed x_i itself.

The two formulae (7.11) and (7.12) can be expressed in the single formula

$$(7.14) \qquad m'_{ij} = m_{ij}(1 - \lambda_i e_{ij})$$

By the definition (7.13) λ_i must be a non-negative quantity, that is

$$(7.15) \qquad 0 \leq \lambda_i.$$

The observed variance m_{ii} must therefore be larger than the systematic variance if assumption (III) is satisfied. The same

would however be true even on a less rigorous assumption. We only need to assume:

(III') There is not such a higher *inverse* correlation between the accidental part of x_i and x_i itself that the third term in (7.10) outweighs the second.

If this is the case, we still have the equation (7.12) and the inequality (7.15), the only difference is that λ_i will now be defined by

(7.16) $$\lambda_i = \frac{m_{ii}'' + 2[x_{ii}' x_{ii}'']}{m_{ii}}.$$

In the following we shall as a rule assume (III') instead of (III).

Assumptions of the kind we have expressed in (I) — (III) must of course be interpreted *cum grano salis*. In practice we must reckon with the possibility that they are not exactly, but only approximately fulfilled.

It may be noted that the assumption (III) does not exclude the possibility of the *size* of the disturbance changing systematically with the size of the variate itself. III only involves that there is such a mixture of positive and negative disturbances that in the total result no correlation appears between x_i and its disturbance.

Let us normalise the variates by dividing each variate by its observed sumsquare. The systematic and accidental parts of each variate must consequently also be conceived of as divided by this magnitude. The moments of the systematic variates thus normalised will be $r_{ij}(1 - \lambda_i e_{ij})$ which may also be written

(7.17) $r_{ij} - \lambda_i e_{ij} =$ moments of systematic parts of empirically normalised variates.

In other words the "true" moment matrix of the normalised variates will be

(7.18) $$\begin{pmatrix} 1 - \lambda_1 & r_{12} \ldots r_{1n} \\ r_{21} & 1 - \lambda_2 \ldots r_{2n} \\ \cdots\cdots\cdots\cdots\cdots \\ r_{n1} & r_{n2} \ldots 1 - \lambda_n \end{pmatrix}$$

where $\lambda_1, \lambda_2 \ldots$ are the disturbing intensities defined by (7.16).

Since the true correlation coefficients are obtained by normalising the true moment matrix, we can also write

$$(7.19) \qquad r_{ij} = \sqrt{(1-\lambda_i)(1-\lambda_j)} \, \varrho_i \qquad (i \neq j)$$

where ϱ_{ij} are the true correlation coefficients and r_{ij} the observed. This equation follows of course also directly from (7.11) and (7.12). If we insert the expression (7.19) for r_{ij} in (7.17) or (7.18), we see that the "true" moments of the (empirically) normalised variates may also be written

$$(7.20) \qquad \varrho_{ij} \sqrt{(1-\lambda_i)(1-\lambda_j)}$$

Here the true moments of the (empirically) normalised variates are expressed exclusively in terms of the true correlations and the disturbances.

8. THE SCATTER FUNCTIONS AND THEIR TAYLOR EXPANSION. THE CUSHION EFFECT.

The discussion of the last Section suggest that the function of $\lambda_1 \ldots \lambda_n$ obtained by taking the determinant of (7.18), and also the similar functions for any given subset $(ij \ldots k)$ will play an important role in any analysis of confluency of regression equations. These functions we shall call the *scatter functions*. They will be denoted $F(\lambda_1 \ldots \lambda_n)$ for the big set and $F_{ij \ldots k}(\lambda_i, \lambda_j \ldots \lambda_k)$ for the subset $(ij \ldots k)$, so that

$$(8.1) \qquad F_{ij \ldots k}(\lambda_i, \lambda_j \ldots \lambda_k) = \begin{vmatrix} 1-\lambda_i & r_{ij} \ldots r_{ik} \\ r_{ji} & 1-\lambda_j \ldots r_{jk} \\ \cdots\cdots\cdots\cdots\cdots \\ r_{ki} & r_{kj} \ldots 1-\lambda_k \end{vmatrix} =$$

$$= (1-\lambda_i) \cdot (1-\lambda_j) \ldots (1-\lambda_k) \begin{vmatrix} 1 & \varrho_{ij} \cdots \varrho_{ik} \\ \varrho_{ji} & 1 \cdots \varrho_{jk} \\ \cdots\cdots\cdots \\ \varrho_{ki} & \varrho_{kj} \cdots 1 \end{vmatrix}$$

If the λ's are put equal to the (so far unknown) disturbing intensities of the variates, the functions $F_{ij \ldots k}$ simply become the "true" scatterances. The second expression in (8.1) is obtained by inserting in the middle expression of (8.1) the expression for r_{ij} taken from (7.19).

Consider the ν-dimensional subset $(ij\ldots k)$, and let $\overline{\lambda}_i, \overline{\lambda}_j \ldots \overline{\lambda}_k$ be any set of given values of the arguments λ. The Taylor expansion of (8.1) is easily found to be

$$(8.2) \quad F_{ij\ldots k}(\lambda_i, \lambda_j \ldots \lambda_k) = \overline{F}_{ij\ldots k} - \sum_\alpha (\lambda_\alpha - \overline{\lambda}_\alpha) \overline{F}_{ij\ldots)\alpha(\ldots k} +$$
$$+ \sum_{\alpha < \beta} (\lambda_\alpha - \overline{\lambda}_\alpha)(\lambda_\beta - \overline{\lambda}_\beta) \overline{F}_{ij\ldots)\alpha\beta(\ldots k} -$$
$$- \ldots + (-)^\nu (\lambda_i - \overline{\lambda}_i)(\lambda_j - \overline{\lambda}_j) \ldots (\lambda_k - \overline{\lambda}_k).$$

here α in $\sum_\alpha$ runs through all the ν affixes $ij\ldots k$. $(\alpha\beta)$ in $\sum_{\alpha<\beta}$ runs through combinations without repetition of the two affixes α and β picked in the set $ij\ldots k$, etc. $\overline{F}_{ij\ldots k}$ denotes the value of $F_{ij\ldots k}$ when the λ's are put equal to the $\overline{\lambda}$'s; and the inverted parentheses denote "exclusion of".

Putting all the $\overline{\lambda} = 0$ we get the expansion of the "true" scatterances in terms of the observed, namely

$$(8.3) \quad F_{ij\ldots k} = \triangle_{ij\ldots k} - \sum_\alpha \lambda_\alpha \triangle_{ij\ldots)\alpha(\ldots k} + \sum_{\alpha<\beta} \lambda_\alpha \lambda_\beta \triangle_{ij\ldots)\alpha\beta(\ldots k} -$$
$$- \ldots + (-)^\nu \lambda_i \lambda_j \ldots \lambda_k$$

and putting $\lambda = 0$ replacing $\overline{\lambda}$ by λ, we get the expansion of the observed scatterances in terms of the "true", namely

$$(8.4) \quad \triangle_{ij\ldots k} = F_{ij\ldots k} + \sum_\alpha \lambda_\alpha F_{ij\ldots)\alpha(\ldots k} + \sum_{\alpha<\beta} \lambda_\alpha \lambda_\beta F_{ij\ldots)\alpha\beta(\ldots k}$$
$$+ \ldots + \lambda_i \lambda_j \ldots \lambda_k.$$

From (8.3) we obtain, of course, immediately the formula (2.2) for the characteristic polynomial by putting all the λ's equal.

Since all the λ are non-negative, and the true scatterances are positive definite determinants, (8.4) shows immediately that an observed scatterance must always (when the conditions I—III of Section 7 hold good) be larger than the corresponding "true" scatterance. In other words, even though the true scatterance $F_{ij\ldots k}$ is rigorously zero, any amount of accidental disturbance will immediately act as a sort of a "cushion" that prevents the observed scatterance from falling down to zero. The thickness of the cushion depends on the intensity of the disturbances. To a first approximation, when only linear terms in the disturbing intensities are retained, the cushion-effect is by

(8.3) proportional to the weighted sum of the intensities, the weights being the observed scatterances in the first subsets.

A similar remark applies to the characteristic roots. Any sort of random disturbances will for instance immediately produce a systematic bias in the direction of *increasing* the minimal λ-root. Indeed, if all the λ_k are replaced by λ and all the $\bar{\lambda}_k$ by $\lambda_k + \lambda$ in (8.2) we get

$$(8.5) \qquad P(\lambda) = F(\lambda) + \sum_\alpha \lambda_\alpha F_{)\alpha(}(\lambda) + \sum_{\alpha < \beta} \lambda_\alpha \lambda_\beta F_{)\alpha\beta(}(\lambda) + \ldots + \lambda_i \lambda_j \ldots \lambda_k$$

where $P(\lambda)$ is the observed and $F(\lambda)$ the "true" characteristic polynomial in the set $(ij \ldots k)$, $F_{)\alpha(}(\lambda)$ the true characteristic polynomial in the subset where α is left out, etc. λ_α etc. in (8.5) are the disturbing intensities. Since the true characteristic polynomial, when written in determinant form, is positive definite for all values of λ not larger than the true minimal root, the development (8.5) shows that as long as we are in the interval between zero and the smallest of the "true" characteristic roots (the latter limit included) we must have $P(\lambda) > F(\lambda)$, unless all the disturbing intensities are zero, in which case $P(\lambda) = F(\lambda)$. Consequently the observed minimal root must be larger than the true, whenever disturbances are present.

More generally the whole appearance of the observed scatter will be influenced by the cushion effect. The shape of the scatter is indeed produced *jointly* by two different sets of factors. It depends, not only on the "true" relations that exist between the systematic variates, but it is systematically biassed by the disturbances. This is just why it is so difficult to apply scatterances or characteristic roots as criteria of linear confluency. And this is also the reason why the shape of the regression ellipsoid (the direction and length of its axes) is not a good indicator of those traits of the distribution which we try to lay bare when we study linear confluency. The things which we take as our landmarks when we study the main axes of the ellipsoid are indeed produced just as much — or perhaps even more — by the cushion effect as by the systematic connection between the variates. This applies particularly to those cases where there is multiple collinearity between the systematic variates, and where consequently a confluence analysis is particularly needed. In this case the direction and length of some of the axes of the regression ellipsoid will depend

primarily on the relative size of the disturbing intensities $\lambda_1 \ldots \lambda_n$ and not on the nature of the "true" relations that exist between the variates.

9. THE ›TRUE‹ REGRESSIONS.

Suppose that we have a set of variates, say Nos. $(ij \ldots k)$ and that we actually *know* the disturbing intensities in the set. Suppose that these intensities are such that they actually make the scatter function $F_{ij \ldots k}$ vanish, but leaving at least one of its first principal minors different from zero, and of course making it a positive definite matrix, otherwise it could not be the moment matrix for the "true" variates. The systematic parts of the observed variates would then be connected by a linear relation, and we could actually determine this linear relation exactly. Indeed, if the "true" regression is written in the form

$$(9.0) \qquad a_1 x_1 + \ldots + a_n x_n = 0$$

or in normal coordinates

$$(9.1) \qquad \alpha_1 \xi_1 + \ldots + \alpha_n \xi_n = 0$$

where ξ_i stand for the empirically normalised variates

$$(9.2) \qquad \xi_i = \frac{x_i}{\sigma_i}, \quad \sigma_i = \sqrt{m_{ii}/N}$$

σ_i being the observed standard deviation of the variate $No \cdot i$, — then the coefficients α_i would simply be the solution of the homogeneous system

$$(9.3) \qquad \sum_i (r_{ik} - \lambda_i e_{ik}) \alpha_k = 0. \qquad (i = 1, 2 \ldots n)$$

In other words the α's could simply be put proportional to the elements in a row of the adjoint of the determinant $|F_{ij \ldots k}|$. All these rows would be proportional since $|F_{ij \ldots k}| = 0$. And at least one of the rows would not consist exclusively of zeros (since we have assumed at least one of its first minors to be different from zero). Incidentally the absolute values of the regression coefficients could also be determined by the square values of the diagonal elements in the adjoint (which are all non-negative according to our assumptions).

If the disturbing intensities are not known exactly, but it is plausible to assume certain *intervals* in which they must lie, we can determine certain *limits* for the "true" regression coefficients.

Let us take the case $n = 2$ as an example. We assume that the observed correlation matrix is given. In the case $n = 2$, all the information contained in this matrix is simply the value of the correlation coefficient r_{12} between the two variates x_1 and x_2. Already the fact that we have given the value of r_{12} imposes a restriction on the possible hypotheses regarding the size of the disturbing intensities λ_1 and λ_2. Indeed, not all such hypotheses are compatible with the observed value of r_{12}. Which hypotheses are so compatible? Obviously those, and only those, that make the "true" moment matrix, which would result from the hypothesis, a positive definite matrix. This means that we have

$$(9.4) \qquad \lambda_1 < 1 \qquad \lambda_2 < 1$$

$$(9.5) \qquad (1 - \lambda_1)(1 - \lambda_2) > r_{12}^2.$$

Graphically these conditions may be exhibited by saying that the point $(\lambda_1 \lambda_2)$ must lie in the area in Figure 7 below and to the right of the two straight lines defined by (9.4), and at the same time below the equilateral hyperbola defined by (9.5) (In Figure 7 $r_{ij} = 0.40$). Obviously the latter condition is sharper than the former, so we may disregard (9.4) and only consider (9.5).

Fig. 7.

Of more interest than (9.4) are the conditions

$$(9.6) \qquad 0 \leqq \lambda_1 \qquad 0 \leqq \lambda_2$$
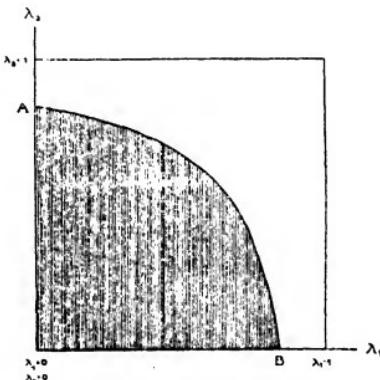
that give *lower* limits for the intensities. This condition is essentially connected with the assumption (III) or (III') of Section 7; (9.6) together with (9.5) define the shaded area in

Figure 7, as the *possibility region* for $(\lambda_1 \lambda_2)$. The larger the square correlation r_{12}^2 the smaller this shaded area. As r_{12}^2 approaches unity the area reduces to a point, namely origin, which means that in the case of perfect observed correlation, there is only one hypothesis possible, namely that no disturbance is present. On the other hand if r_{12}^2 approaches zero, the possibility region will at the limit fill the whole square of Figure 1.

The condition (9.5) in conjunction with (9.6) entails

$$(9.7) \qquad \lambda_1 < 1 - r_{12}^2 \qquad \lambda_2 < 1 - r_{12}^2$$

These latter inequalities express an *independent* upper limit for each disturbing intensity in terms of observed parameters, but of course this limit is in general not so sharp as the joint limit (9.5).

If we supplement the information contained in the observed r_{12} with the assumption that the systematic variation of the two variates are linearly dependent, the possibility region is of course further limited. In this case we must have

$$(9.8) \qquad \begin{vmatrix} 1 - \lambda_1 & r_{12} \\ r_{21} & 1 - \lambda_2 \end{vmatrix} = 0$$

so that the point $(\lambda_1 \lambda_2)$ must now lie on the limiting hyperbola itself; in other words there is now only a one dimensional freedom in our choice of hypotheses about $(\lambda_1 \lambda_2)$. *This entails definite limits for the "true" regression coefficients.* Indeed, if λ_1 and λ_2 are so chosen that (9.8) is fulfilled, the regression coefficients are given by any of the rows of the adjoint of (9.8), that is by any row in

$$(9.9) \qquad \begin{vmatrix} 1 - \lambda_2 & - r_{12} \\ - r_{21} & 1 - \lambda_1 \end{vmatrix}$$

We may for instance put

$$(9.10) \qquad \xi_1 = \frac{r_{12}}{1 - \lambda_2} \xi_2$$

where $\xi_1 \xi_2$ are the normalised observed variates. What are the limits of variation for the regression coefficient $\dfrac{r_{12}}{1 - \lambda_2}$ when the point $(\lambda_1 \lambda_2)$ varies on the possibility hyperbola of Figure 1?

Obviously this regression coefficient becomes the largest or smallest in absolute value in the points A and B respectively. In these two extremum points we have $1 - \lambda_2 = r_{12}^2$ and $1 - \lambda_2 = 1$. Hence: The "true" regression coefficient between the normalised variates ξ_1 and ξ_2 must lie between r_{12} and $\dfrac{1}{r_{12}}$. In other words:

(9. 11)   *The "true" linear regression between two variates must lie between the two elementary regressions.*

This result is deduced on the assumption that there actually exist a linear relation between the *systematic* parts of the two variates and further that (I), (II) and (III') of Section 7 are fulfilled.

This suggests that similar limits may be obtained by study- ing in several variates the condition that the scatter functions (8. 1) shall be positive definite. The analogue of (9. 7) is for instance found to be

$$(9.\ 12) \qquad\qquad \lambda_\alpha \lessgtr \frac{\Delta_{ij\ldots k}}{\Delta_{ij\ldots)\alpha(\ldots k}}$$

where $(ij \ldots k)$ is any set containing α. The limit (9. 12) is all the narrower the more inclusive the set $(ij \ldots k)$.

So far as the regression coefficients themselves are concern- ed, the situation becomes however much more complex in several variates, particularly because it may now be possible to find points ("subtest points") where, not only the highest order scatter function vanishes, but also all those of next lower order. At the Institute considerable work, both theoretical and numeri- cal, has been devoted to clearing up this matter in the general case of n variates, but the results are not yet in such a shape as to justify publication. I shall therefore here confine myself to the above example in two variates. This example will be sufficient to suggest heuristically one of the leading ideas which will be utilised in Part III.

It was explained above how the "true" regression in any number of variates can be determined, when the conditions of Section 7 hold good and the disturbing intensities are assumed given. The regression coefficients are then simply proportional to the elements in the adjoint of the scatter function. The determination of the "true" regression can also be thrown into another form which is independent of the conditions of Section

7 and which brings forth another set of parameters characteris·
ing the nature of the disturbances.

If the observed variates consist of two parts as indicated in
(7.8), and if the there exist a linear relation between the
systematic parts of the variates, we must have

$$(9.13) \qquad a_1 x_1 + \ldots + a_n x_n = u$$

where the a_k are constants (the "true" regression coefficients),
x_k the observed variates, and u a variate that may be looked
upon as a *shift* of the regression plane. In terms of the
disturbances x_k'' of the individual variates the shift u may be
expressed as

$$(9.14) \qquad u = a_1 x_1'' + \ldots + a_n x_n''.$$

For a moment let us disregard the composition of u as defined
by (9.14) and let us just consider it as a variate defining the
shift of the equation (9.13). Multiplying (9.13) by x_i and
performing a summation over all the observations, we get

$$(9.15) \qquad \sum_k m_{ik} a_k = [ux_i].$$

For the coefficients α of the normalised regression equation
(that is of (9.1)) we have

$$(9.16) \qquad \alpha_k = a_k \sigma_k \qquad \qquad \sigma_k = \sqrt{m_{kk}/N}$$

consequently

$$(9.17) \qquad \sum_k r_{ik} \alpha_k = \varepsilon \cdot \varrho_i$$

where

$$(9.18) \qquad \varepsilon = \sqrt{[uu]/N}$$

is the standard deviation of the shift, and

$$(9.19) \qquad \varrho_i = \frac{[ux_i]}{\sqrt{[uu][x_i x_i]}}$$

the correlation coefficient between the observed variate x_i and
the shift.

The equation (9.17) is perfectly general, and so far does not

depend on any assumption about the shift. If the disturbance correlations ϱ_i are *given*, then (9.17) permits a unique determination of the regression coefficients. Indeed — the determinant of the system (9.17) is the actually observed correlation matrix (r_{ij}) which will never be zero in practice. The fact that the standard deviation ε of the shift occurs in the right member of (9.17) is unimportant; indeed, in the solution this parameter will only appear as a common *multiplier* for all the regression coefficients, and the regression equation is independent of such a factor. We may therefore write the solution of (9.17) in the form

$$(9.20) \qquad \alpha_i = \text{constant} \sum_k \hat{r}_{ik}\,\varrho_k$$

Thus, what the observational data at hand gives information about is in reality not the true regression coefficients themselves but only *the law of linear transformation which permits to pass from the disturbance correlations ϱ_i to the regression coefficients α_i or vice versa.* This exhibits in a striking fashion the limits of the information which are contained in a table of correlation coefficients. From such a table we can deduce some definite statistical regression only by a process which is equivalent to making certain hypotheses about the disturbance correlations. We shall later interpret some of the usual statistical regressions in this light.

The hypothesis that the shift is uncorrelated with all the observed variates i·e·$\varrho_1 = \ldots = \varrho_n = 0$, can in practice not be admitted Indeed, this would — if we disregard the trivial case where all the true regression coefficients are zero — entail the vanishing of the *observed* correlation determinant. This follows immediately from the fact that (r_{ij}) is the matrix of the linear system (9.17).

The determination of the "true" regression coefficients obtained by means of a given set of assumptions regarding the ϱ must of course be identical with the determination by means of an assumption about the λ's of Section 7. Suppose for simplicity that we admit the three conditions (I), (II), (III) of Section 7. The moment $[u\,x_i]$ is then equal to

$$(9.21) \qquad [u\,x_i] = \sum_k a_k\,[x_k^*\,x_i] = a_i\,\lambda_i\,m_{ii}.$$

The equation (9.15) therefore takes on the form

$$(9.22) \qquad \sum_k (m_{ik} - \lambda_i \, m_{ii} \, e_{ik}) \, a_k = 0$$

which is equivalent to (9.3).

In non-normalised form the "true" regression defined by the coefficients (9.20) may be written

$$(9.23) \qquad \sum_{ij} \mu_i \hat{m}_{ij} x_j = 0$$

were μ_i is the disturbance moment

$$(9.24) \qquad \mu_i = [u \, x_i].$$

In determinant form this "true" regression may be written

$$(9.25) \qquad \begin{vmatrix} 0 & x_1 \ldots x_n \\ \mu_1 & m_{11} \ldots m_{1n} \\ \cdots\cdots\cdots\cdots \\ \mu_n & m_{n1} \ldots m_{nn} \end{vmatrix} = 0.$$

This regression is — as one would expect — *invariant* for a general linear transformation of the observational variates. Indeed, making the transformation

$$(9.26) \qquad \overline{x}_i = \sum_k c_{ik} x_k$$

where (c_{ik}) is a non-singular matrix, we get a new observational moment matrix

$$(9.27) \qquad \overline{m}_{ik} = \sum_{hk} c_{ik} m_{kh} \partial_{hj} \text{ where } \partial_{hj} = c_{jh}$$

(Compare formula (1.5) in "Correlation and Scatter..."). The adjoint of (9.27) is

$$(9.28) \qquad \overset{\blacktriangle}{\overline{m}}_{ij} = \sum_{hk} \check{c}_{ih} \, \hat{m}_{hk} \, \hat{c}_{kj}$$

$\check{c}$ being the adjoint of ∂, and $\hat{c}$ the adjoint of c.

In each observation the shift u will be unchanged by the linear transformation; the transformation represents indeed only another way of arranging the terms in the sum to the left in (9.13), and this purely formal operation cannot influence the value which "nature" has given this sum. The new disturbance moment will consequently be

$$(9.29) \qquad \overline{\mu}_i = [u\overline{x}_i] = \sum_k c_{ik}\mu_k.$$

In other words μ_i is cogredient with x_i, which is the essential fact that will ensure the invariance of the regression. The new regression equation will now be $\sum_{ij} \overline{\mu}_i \hat{m}_{ij} \overline{x}_j = 0$, and inserting here from (9.26), (9.28) and (9.29) we get

$$\sum_{ij} \sum_{hk} \sum_{\alpha\beta} \mu_\alpha \check{c}_{\alpha i} \check{c}_{ih} \hat{m}_{hk} \hat{c}_{kj} c_{j\beta} x_\beta = 0$$

which reduces to

9(.30) $$|c|^2 \cdot \sum_{\alpha\beta} \mu_\alpha \hat{m}_{\alpha\beta} x_\beta = 0.$$

Since the determinant $|c|^2$ is different from zero, equations (9.30) and (9.23) are the same.

In (9.11) we deduced limits for the "true" regression coefficient in two variates by discussing the range of variation of the λ's. Will a similar study of the range of variation of the $\varrho's$ in (9.20) also furnish limits for the "true" regression coefficients? For simplicity let us again take the case $n = 2$. By (9.20) the regression equation in normalised coordinates will now be

(9.31) $$\xi_1 = \beta \cdot \xi_2$$

where

(9.32) $$\beta = \beta(z) = \frac{1 - zr}{r - z}, \qquad r = r_{12}.$$

(9.33) $$z = \varrho_1/\varrho_2.$$

The function $\beta(z)$, whose derivative is $\beta'(z) = \dfrac{1 - r^2}{(r - z)^2}$ will have one branch that starts at $\beta = r$ for $z = -\infty$, increasing monotonically to $+\infty$ for $z = r - 0$, and appearing again at $-\infty$ for $z = r + 0$, from where it increases monotonically to r. When we follow these branches, we see that β lies between r and $1/r$ when *and only when z has the opposite sign of r.* Consequently: If the disturbing correlations ϱ_1 and ϱ_2 have the *same* sign when the observed r is negative and if they have *different* signs when r is positive, then, and only then, will the "true" regression lie between the two elementary regressions. If this condition is not fulfilled, the "true" regression will fall outside of the sector between the elementary regression. In this latter case there

is nothing to prevent the "true" regression slope to assume any value between $-\infty$ and $+\infty$. This is another expression for the fact that the limitation (9.11) is essentially connected with the special assumptions (I), (II) and (III') of Section 7.

10. THE INTERPRETATION OF THE EMPIRICAL REGRESSIONS: ELEMENTARY, ORTHOGONAL, DUO-ORTHOGONAL AND DIAGONAL REGRESSIONS.

By the formulae of the preceding Section we may now study the nature of the various empirical regressions. Each of the usual kinds of regressions determined statistically is connected with a definite kind of assumption about the disturbing intensities λ or the disturbing correlations ϱ.

Let us first take the elementary regressions. If we adopt the p-th elementary regression as an expression for the systematic connection between the variates, we assume that the coefficients of the "true" regression (9.23) are proportional to those of the p-th line in the adjoint of the observed moment matrix; in other words, we assume that

$$(10.1) \qquad \sum_i \mu_i \hat{m}_{ij} = C_p \hat{m}_{pj} \qquad (j = 1, 2 \ldots n)$$

where C_p is some constant independent of j.

From (10.1) follows immediately — provided that the observed moment matrix is non-singular — that

$$(10.2) \qquad \mu_i = C_p\, e_{ip} = \begin{cases} 0 & (i \neq p) \\ C_p & (i = p) \end{cases}$$

In other words, adopting the p-th elementary regression involves necessarily the assumption that there is some correlation between the shift u and the observed variate x_p, but no correlation whatsoever between u and any of the other observed variates.

Next consider the orthogonal regression. It is defined as the regression obtained by minimising the sumsquare of the deviations measured perpendicularly to the regression plane. If this is done before the variates are normalised, we obtain a regression that is not even invariant for a change in units of measurement. If the orthogonal regression is to be used at all in a case where the units of measurements are conventional, it should therefore be applied to the normalised variates. If

this is done, the regression coefficients α are nothing but the solution of (9 3) when all the λ's are put equal, and equal to the smallest characteristic root, i. e. the smallest zero of $P(\lambda)$ defined by (2.2). In other words we may look upon the normalised orthogonal regression as obtained by first assuming that the disturbing intensity is the same for all the variates, and then determine this common magnitude of the disturbing intensity as the smallest number compatible with the assumption that the *systematic* parts of the variates are rigorously linearly dependent.

This suggests the generalisation of estimating — through some more or less plausible considerations of the concrete nature of the variates — a set of proportionality numbers $\lambda_1' \ldots \lambda_n'$ which can roughly express the *comparative* amount of disturbance on the variates; and then determine a common factor γ by the characteristic equation $F(\gamma\lambda_1' \ldots \gamma\lambda_n') = 0$. The development of this equation is easily obtained from the formulae of the preceding Sections.

When we are primarily interested in the regression coefficient between two special variates x_p and x_q (the others being taken into the regression only to eliminate influences that it is not wanted to study), we may for instance use the above method by putting all the λ' equal to zero except λ_p' and λ_q', which we may — if no further information is available — put equal. The regression thus obtained may be called the *duo-orthogonal*. The explicit formula for its computation is

(10. 3)

$$B_{pq(\alpha,\beta\ldots\gamma)}^{(\text{duo})} = \varepsilon \sqrt{\frac{2\,\Delta_{\alpha\beta\ldots)q(\ldots\gamma} - \Lambda}{2\,\Delta_{\alpha\beta\ldots)p(\ldots\gamma} - \Lambda}} = \varepsilon \sqrt{\frac{\sqrt{U} - (\Delta_{)p(} - \Delta_{)q(})}{\sqrt{U} + (\Delta_{)p(} - \Delta_{)q(})}}$$

where ε is a sign factor determined from the general appearance of the adjoint correlation matrix, and

(10. 4 a)

$$\Lambda = (\Delta_{\alpha\beta\ldots)p(\ldots\gamma} + \Delta_{\alpha\beta\ldots)q(\ldots\gamma})$$

$$- \sqrt{(\Delta_{\alpha\beta\ldots)p(\ldots\gamma} + \Delta_{\alpha\beta\ldots)q(\ldots\gamma})^2 - 4\,\Delta_{\alpha\beta\ldots\gamma}\,\Delta_{\alpha\beta\ldots)pq(\ldots\gamma}}$$

(10. 4 b) $\qquad U = (\Delta_{)p(} + \Delta_{)q(})^2 - 4 \cdot \Delta \cdot \Delta_{)pq(}$

The last member in (10. 3) is most convenient for the numerical computation.

Also the elementary regressions can be interpreted in terms of the λ's. Indeed suppose that there is no disturbance at all on the variates except on x_p. The true scatterance in the given set will then be the value of (8.1) when all the λ's are just equal to zero except λ_p. The value which λ_p must have if the systematic variates in the set shall be linearly dependent, that is $F_{ij} \ldots k = 0$, will consequently be

$$(10.5) \qquad \lambda_p = \frac{\Delta_{ij \ldots k}}{\Delta_{ij \ldots)p(\ldots k}}$$

Any row in the adjoint of the scatter function with the value (10.5) inserted for λ_p will then give the regression coefficients. The elements in the p-th row of the adjoint thus obtained will just turn out to be the coefficients of the p-th elementary regression.

The *diagonal* regression is the one obtained by determining the signs of the regression coefficients from an inspection of the signs of the elements in the adjoint correlation matrix, and then putting the absolute value of the coefficients equal to the square root of the diagonal elements in the adjoint. Obviously this regression is the one obtained by assuming that the disturbing intensities are such that the "cushion" effect (i.e. the effect that prevents the observed scatterances in the $(\nu-1)$ dimensional subsets from coming down to the value of the "true" scatterances) is the same in all subsets. In other words, the observed scatterances in these subsets are assumed proportional to the true.

The diagonal regression may be computed by the formula

$$(10.6) \qquad B_{pq(\alpha, \beta \ldots \gamma)}^{(\text{diag.})} = \varepsilon \sqrt{\frac{\hat{r}_{qq\ (\alpha\beta \ldots \gamma)}}{\hat{r}_{pp\ (\alpha\beta \ldots \gamma)}}} = \varepsilon \sqrt{\frac{\Delta_{\alpha\beta \ldots)q(\ldots \gamma}}{\Delta_{\alpha\beta \ldots)p(\ldots \gamma}}}$$

where ε is the sign factor indicated.

In the above analysis the regressions are not determined by any least square minimalisation procedure, but simply by specifying certain *assumptions* about the disturbing intensities or the disturbing correlations and then solving the equations that must exist *exactly* if the assumptions in question hold good. It seems that this is a more logical procedure than just to least square minimise certain deviations defined in a more or less empirical manner. If a least square process is to be taken as

the basis for the determination of the regression coefficients, it should at least be formulated in more general terms than those usually employed. It should be formulated so that the nature of the specialisation adopted in each particular case is clearly exhibited. The following is a suggestion for such a treatment of the problem.

Let (9.0) be the empirical regression plane. Suppose that the a's are to be determined by minimising the sumsquares of the deviations from the plane measured in a direction whose direction numbers are $c_1 \ldots c_n$. These direction numbers may be a set of given constants, or more generally they may be given *functions* of the a's. The nature of these functions are just characteristic for the nature of the minimalisation process considered. We shall treat the problem on the assumption that the minimalisation is made in the original (non-normalised) variates. The corresponding solution obtained in the normalised variates is obtained simply by assuming all the observed standard deviations (or sumsquares) to be unity.

Let x_i be the coordinates of a given observation point and v_i the coordinates of the point of intersection between the regression plane and the straight line through x_i with direction c_i. We then have

$$(10.7) \qquad v_i - x_i = -\frac{\sum_k a_k x_k}{\sum_k a_k c_k} c_i$$

so that the square distance is

$$(10.8) \qquad \sum_k (v_k - x_k)^2 = (\sum_k a_k x_k)^2 \cdot \sum_k c_k^2 / (\sum_k a_k c_k)^2.$$

Taking the sum of this over all observations, we get

$$(10.9) \qquad \lambda = \frac{\sum_k c_k^2}{(\sum_k a_k c_k)^2} \cdot \sum_{ij} m_{ij} a_i a_j.$$

Equating the partial derivatives of λ with respect to the a's to zero we get

$$(10.10) \qquad \sum_k m_{ik} a_k = \lambda \mu [c_i + \sum_k a_k c_{ki} - \mu \sum_k c_k c_{ki}]$$

where

$$(10.11) \qquad \mu = \sum_k a_k c_k / \sum_k c_k^2,$$

(10. 12)
$$c_{ki} = \frac{\delta c_k}{\delta a_i}$$

It does not restrict generality if we assume the c_k to be such functions of the a_k that

(10. 13)
$$c_1^2 + \ldots + c_n^2 = 1$$

for all values of the a_k.

If this is the case, (10. 10) reduces to

(10. 14)
$$\sum_k m_{ik} a_k = \lambda \mu (c_i + \sum_k a_k c_{ki}) \quad (i = 1, 2 \ldots n).$$

In this system the c_i and c_{ki} must be supposed to be given functions of the a_k. In general the system (10. 14) will therefore furnish a determination of the a_k.

If we assume that the minimalisation is done in a direction that is independent of the inclination of the regression plane, (10. 14) reduces further to

(10. 15)
$$\sum_k m_{ik} a_k = \lambda \mu c_i$$

where now $\lambda \mu$ may be looked upon as an arbitrary parameter, which only defines the *length* of the vector a_k. The regression equation is of course independent of this length.

If all the c_i in (10. 15) are put equal to zero, except one, we get the corresponding elementary regression.

Putting in (10. 14)

(10. 16)
$$c_i = \frac{a_i}{\sqrt{\sum_k a_k^2}}$$

means that we consider the orthogonal regression. In this case we get

(10. 17)
$$c_{ij} = \left(e_{ij} - \frac{a_i a_j}{\sum_k a_k^2} \right) \Big/ \sqrt{\sum_k a_k^2}.$$

Also in this case we therefore have

(10. 18)
$$\sum_k a_k c_{ki} = 0. \qquad (i = 1, 2 \ldots n)$$

so that we get back to equation (10. 15).

Further, we now have $\mu c_i = a_i$, so that (10. 15) reduces to

(10. 19) $$\sum_k (m_{ik} - \lambda e_{ik}) a_k = 0$$

which is the previously encountered equation for determining the coefficients of the orthogonal regression.

Any number of other regressions may be defined by specifying the nature of the functions c_i. A rather extensive and interesting class is obtained by considering all functions of $a_1 \ldots a_n$ for which (10. 18) is fulfilled.

For all functions of this class the problem will be reduced to a system of the form (10. 15). If further the functions c_i are of the form

(10. 20) $$c_i = \varepsilon_i a_i \varPhi$$

where ε_i is a given constant depending on i but independent of $a_1 \ldots a_n$ and $\varPhi$ a function of $a_1 \ldots a_n$ and $c_1 \ldots c_n$, but independent of i, then the problem reduces to the solution of a regular characteristic equation of the form

(10. 21) $$|m_{ij} - \varLambda \varepsilon_i e_{ij}| = 0$$

$\varLambda$ being the unknovn. Indeed, in this case (10. 15) reduces to the system

(10. 22) $$\sum_k (m_{ik} - \varLambda \varepsilon_i e_{ik}) a_k = 0, \text{ where } \varLambda = \lambda \mu \varPhi$$

The system (10. 22) has a non-trivial solution only when $\varLambda$ satisfies (10. 21).

An example of a regression that comes under the above class is the one obtained by constructing through each observation point a plane P perpendicular to a certain set of the coordinate axes and then minimising the sumsquare of the deviations measured within P and perpendicularly to the manifold of intersection between P and the regression plane. This regression may perhaps be called the subset-orthogonal. The duo-orthogonal mentioned above is a special case.

Many other more or less plausible procedure may be derived by other specialisations of the functions c_i. However, if a logical interpretation of all these various procedures is wanted it seems that we must bring the problem back to a discussion of the effects of the parameters λ and ϱ along the lines indicated in the beginning of this Section.

11. THE AMOUNT OF INDETERMINATENESS IN REGRESSION SLOPES IN MULTICOLLINEAR VARIATES. THE PERSISTENCY EFFECT PRODUCED WHEN A GIVEN INTER-COEFFICIENT HAS THE SAME SIZE IN ALL SUBSETS.

Utilising the results of the preceding Section we shall now study more closely the amount of indeterminateness in regression slopes in multicollinear variates.

Suppose first that we have a set of n variates $x_1 \ldots x_n$ between which there exist the two independent and exact linear relations

$$(11.1) \qquad a_1 x + \ldots + a_n x_n = 0$$

$$(11.2) \qquad b_1 x_1 + \ldots + b_n x_n = 0.$$

Let p and q be two arbitrary multipliers and let us take the sum of (11.1) and (11.2) multiplied by p and q respectively; this gives the new equation

$$(11.3) \qquad c_1 x_1 + \ldots + c_n x_n = 0$$

where

$$(11.4) \qquad c_k = p a_k + q b_k.$$

Since p and q are arbitrary, the coefficients of the new equation contain also an element of arbitrariness. The extent of this arbitrariness is such that in the new equation we may make the regression coefficient between any given set of two variates, say between x_i and x_j equal to any number we like, provided only that

$$(11.5) \qquad \begin{vmatrix} a_i & a_j \\ b_i & b_j \end{vmatrix} \neq 0.$$

The above rule is immediately verified if we write (11.3) in the form

$$(11.6) \qquad x_i = \beta x_j + \ldots$$

The coefficient β of this equation will then be

$$(11.7) \qquad \beta = -\frac{p a_j + q b_j}{p a_i + q b_i}$$

By a suitable choice of p and q the expression (11.7) may under the assumption (11.5) be made equal to any preassigned quantity.

Of course, if the multipliers p and q are chosen so as to make β equal to a given number, the *other* regression coefficients will be determined. In other words, if there exist *two* independent equations of the form (11.1) and (11.2), we are only at liberty to choose *one* of the regression coefficients in the derived equation (11.6), that is to say the regression coefficients in the derived equation have a *one*-dimensional arbitrariness.

More generally, suppose that there exists $\varkappa$ $(\overline{\leq} n)$ independent equations of the form

$$\sum_i a_K\, x_i = 0$$

(11.8)

$$(K = 1, 2 \ldots \varkappa)$$

The condition that these equations shall be independent is equivalent to the condition that the $\varkappa$ rowed and n-columned matrix

(11.9)
$$\left\| \begin{array}{c} a_{11} \ldots a_{1n} \\ \cdots\cdots\cdots \\ a_{\varkappa 1} \ldots a_{\varkappa n} \end{array} \right\|$$

shall be of rank $\varkappa$. If this condition fulfilled, the $\varkappa$ equations considered are always compatible when $\varkappa \overline{\leq} n$. Further, suppose that the $\varkappa$-rowed and two-columned matrix

(11.10)
$$\left\| \begin{array}{cc} a_{1i} & a_{1j} \\ a_{2i} & a_{2j} \\ \cdots\cdots \\ a_{\varkappa i} & a_{\varkappa j} \end{array} \right\|$$

is of rank two, which means that the $\varkappa$ numbers $a_{1i} \ldots a_{\varkappa i}$ are not proportional to the $\varkappa$ numbers $a_{1j} \ldots a_{\varkappa j}$, at least one pair falls out of proportion. Then the regression coefficients between x_i and x_j in the general equation, that now exists between the variates, may be chosen quite arbitrarily. Indeed, from the equations (11.8) we may now deduce an equation of the form (11.3), where

(11.11)
$$c_i = \sum_K p_K\, a_{Ki}$$

p_K being a set of arbitrary multipliers. The regression coefficient of x_i on x_j in this new equation is

$$(11.\,12) \qquad \beta = -\frac{\sum_K p_K a_{Kj}}{\sum_K p_K a_{Ki}}$$

and if the condition specified under (10. 10) is fulfilled, we may by a suitable choice of the p_K make (11. 12) equal to any number we please.

Still more generally, if $(ij \ldots k)$ is a ν-dimensional subset, $\nu \leqq \varkappa \leqq n$, and the $\varkappa$-rowed and ν-columned matrix

$$(11.\,13) \qquad \left\| \begin{array}{c} a_{1i}\, a_{1j} \ldots a_{1k} \\ \cdots\cdots\cdots \\ a_{\varkappa i}\, a_{\varkappa j} \ldots a_{\varkappa k} \end{array} \right\|$$

is of rank ν, then all the ν coefficients $c_i, c_j \ldots c_k$ in the general equation between the variates may be chosen arbitrarily. This is seen by considering

$$(11.\,14) \qquad \sum_K p_K a_{K\alpha} = c_\alpha \qquad (\alpha = i,j \ldots k)$$

as a system of ν equations in the p_K. Retaining only such a set of ν magnitudes p_K as will make the coefficient matrix of (11. 14) non-singular, (and putting all the other p_K equal to zero) we see that a solution in the p_K is possible for any selection of the c_i. If we are only interested in the *proportions* between the regression coefficients, we may now therefore say that the coefficients have a $(\nu-1)$-dimensional arbitrariness. (10. 10) represents the case $\nu = 2$, and at the other extreme we have the case $\nu = n$; in this case the only possible values of the variates are $x_1 = \ldots = x_n = 0$, so that all the regression coefficients, independently of each other, may now be put equal to any values we please.

Of course, in none of the situations discussed above (when $\varkappa > 1$) has it a sense to speak of *the* regression equation connecting the variates, since no such determinate equation exists.

If $\varkappa > 1$, some sort of *side conditions* may be considered which will make the regression coefficients determinate. Consider, for instance, the case where we have four variates $x_1 \ldots x_4$ satisfying two independent relations. Any set of *three* variates will then form a linearly dependent set. If (11. 5) is fulfilled for $i = 1, j = 2$, it will now have no meaning to speak of the regres-

sion coefficient between x_1 and x_2 in the big set (1234). But it will have a meaning to speak of this regression coefficient in any of the three dimensional subsets containing (12), that is in the sets (123) and (124). We can express this in the usual regression coefficient notation by saying that $b_{12.34}$ has no meaning, while $b_{12.3}$ a· d $b_{12.4}$ have meanings. The latter two coefficients can be looked upon as those obtained from (11.12) by imposing in one case the side condition $c_4 = 0$, and in the other the condition $c_3 = 0$, the c's being the coefficients in the general equation (11.3) (which now contains the four variates $x_1 \ldots x_4$).

The one important exception to the above is when (11.5) is not fulfilled. In this case it *will* have a meaning to speak of *the* regression coefficient between x_1 and x_2 without specifying the subset considered, or giving any other kind of side condition. As an example consider the case where x_2 is *lacking* in both equations, that is, we have $a_2 = b_2 = 0$. Now it will have a meaning to speak of the regression coefficient of x_1 on x_2 even in the big multiply collinear set containing all the four variates. Indeed, whatever multipliers we use to obtain the new general equation, the regression coefficient of x_1 on x_2 would in this new equation be zero. More generally this regression coefficient of x_1 on x_2 would be independent of the multipliers, even if x_2 was not lacking in both original equations, provided only that the regression coefficient between x_1 and x_2 was *the same* in both these equations, in other words provided only that $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = 0$. In this case it is indifferent for the definition of the regression coefficient which one of the various derived equations we consider. In particular the coefficient considered is the same no matter which one of the various subsets we consider. We shall refer to this fact by saying that the regression coefficient in question now shows a *persistency* effect even in the multicollinear set.

In the above example we only had two equations ($\varkappa = 2$). In the general case, there will be produced a persistency effect for the regression coefficient between the two special variates x_i and x_j whenever the matrix (11.10) is of rank one. And if the matrix (11.13) is of rank one, there will be produced a persistency effect for the regression coefficient between *any* pair of variates in the set $(ij \ldots k)$.

So much for the situations that arise when the variates fulfil exactly certain linear relations. Now suppose that each of the n variates $x_1 \ldots x_n$ is the sum of a systematic component and a disturbance as explained in Section 7. Suppose that there exist two or more independent linear relations between the systematic parts of these variates, but that we are not aware of this multi-collinearity and proceed to determine an empirical regression equation between all the variates.

If the empirical regression equation is taken in the diagonal form (10. 6) the squares of the regression coefficients — when normalised variates are used — are simply the correlation determinants $\triangle_{\,(}$. It is therefore to be expected that the determination of the regression coefficients in the present case will be particularly strongly influenced by a cushion effect similar to the one we studied in Section 8 for scatterances, minimal roots etc. Let us see what form this will take in the present case. By (8. 4) the determinant $\triangle_{)i(}$ is equal to

$$(11.\ 15) \quad \triangle_{)i(} = F_{)i(} + \sum_\alpha \lambda_\alpha\, F_{)i\alpha(} + \sum_{\alpha<\beta} \lambda_\alpha \lambda_\beta F_{)\,i\alpha\beta(} + \\ + \ldots + \lambda_1 \lambda_2 \ldots) \lambda_i(\ldots \lambda_n.$$

where $F_{)i(},\ F_{)i\alpha(}$, etc. are the "true" scatterances in the various subsets.

Let us first suppose that the systematic parts of the n variates, that form the ig set, are linearly dependent but that there is some $(n-1)$-dimensional subset whose systematic parts are not linearly dependent. This means that at least for some i will $F_{)i(}$ be different from zero. It now has a meaning to speak of *the* regression equation between the systematic parts of the variates, and $F_{)i(}$ is the square of the "true" regression coefficient that occurs in front of x_i when the "true" regression is written in homogeneous form. Further, let us suppose that, not only is, (for some i), $F_{)i(}$ in the right member of (11. 15) different from zero, but that the disturbing intensities λ are small enough to make this $F_{)i(}$ the *principal* term in the right member of (11. 15). In this case it has a meaning to take the *observed* $\triangle_{)i(}$ as an approximation to the square of the "true" regression coefficient.

But if there exists exactly *two* independen equations between the systematic parts of the variates, then all the $F_{)i(}$ will be

zero but not all the $F_{\lambda_i j_i}$. Hence, all the first terms in the right member of (11.15) will vanish, and the principal terms in the empirically determined diagonal regression coefficients will now depend essentially on the disturbing intensities. Any other type of empirical regression (elementary, orthogonal etc.) would show a similar effect.

A similar consideration applies if the systematic parts of the variates are connected by a certain number $\varkappa$ of independent equations. The only difference is that now all the first $(\varkappa - 1)$ terms in the right member of (11.15) will vanish.

To illustrate the above general tendency let us consider three observational variates x_1, x_2, x_3, and let us suppose that there exist a structural relation of the form

$$(11.16) \qquad x_1 = \beta_{12} x_2 + \beta_{13} x_3 + y$$

where β_{12} and β_{13} are "true regression coefficients and y a disturbance. Multiplying (11.16) by x_i $(i = 2,3)$ and extending a summation over all the observations, we get

$$(11.17) \qquad m_{1i} = \beta_{12} m_{2i} + \beta_{13} m_{3i} + [x_i y].$$

The m's are the observed moments and $[x_i y]$ the unobserved disturbance moments. If on the basis of the observations of x_1, x_2, x_3 we determine the elementary regression of x_1 on x_2 and x_3 we get

$$(11.18) \qquad b_{12.3} = \frac{(m_{21} m_{33} - m_{31} m_{23})}{(m_{22} m_{33} - m_{23}^2)}$$

and similarly for $b_{13.2}$. Inserting into (11.18) the expression for m_{12} and m_{13} taken from (11.17) we obtain

$$(11.19) \qquad b_{12.3} = \frac{\begin{vmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{vmatrix} \cdot \beta_{12} + \begin{vmatrix} [x_2 y] & m_{23} \\ [x_2 y] & m_{33} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{vmatrix}}.$$

If x_2 and x_3 are not exactly collinear, (11.19) may be written in the form

$$(11.20) \qquad b_{12.3} = \beta_{12} + \frac{[x_3 x_3] \cdot \sqrt{[yy] \cdot [x_2 x_2]}}{\triangle_{23}} \cdot \begin{vmatrix} \varrho_2 & r_{23} \\ \varrho_3 & 1 \end{vmatrix}$$

where $\triangle_{23}$ is the scatterance in the set (23) and ϱ_i the correlation between the disturbance variate and the observed x_i.

This illustrates that the error term — that is the second term in (11.20) — becomes all the more important the more perfect the correlation between x_2 and x_3. The error term is inversely proportional to the scatterance in the set (23) so that as this scatterance decreases towards zero, the error term will increase beyond any limit, and may therefore easily become *dominating* besides the "true" term β_{12}.

Apart from the trivial coincidence when the determinant to the right in (11.20) vanishes, it is only when there is no disturbance, i. e. when $y = 0$, that the result is independent of the amount of correlation that exist between x_2 and x_3. In this case the elementary regression coefficient $b_{12.3}$ will by (11.20) always be exactly equal to the "true" coefficient β_{12}, provided only that x_2 and x_3 are not *exactly* linearly connected, in which case (11.19) shows that $b_{12.3}$ become of the form $\frac{0}{0}$.

12. THE STABILITY OF A REGRESSION COEFFICIENT UNDER THE INCLUSION OF AN EXTRANEOUS VARIATE.

The fictitious determinateness studied at the end of the last Section is even such that it may exhibit a certain measure of stability as we go from one set of variates to another. As an example suppose that we have three variates x_1, x_2, x_3 between whose systematic parts there exists two independent relations, so that $F_{)1(} = F_{23} = 0$, $F_{)2(} = F_{13} = 0$, $F_{)3(} = F_{12} = 0$. For the lower order true scatterances we have by (8.1) $F_{)12(} = F_3 = 1 - \lambda_3$ etc., so that by (11.15)

$$(12.1) \qquad \begin{aligned} \triangle_{)1(} &= \triangle_{23} = \lambda_2 + \lambda_3 - \lambda_2 \lambda_3 \\ \triangle_{)2(} &= \triangle_{13} = \lambda_1 + \lambda_3 - \lambda_1 \lambda_3 \\ \triangle_{)3(} &= \triangle_{12} = \lambda_1 + \lambda_2 - \lambda_1 \lambda_2 \end{aligned}$$

The square of the regression coefficient of x_1 on x_2 as "determined" by the diagonal regression in the set (123) will consequently be a ratio between erratic elements, namely

$$(12.2) \qquad (B_{123}^{(\text{diag})})^2 = \frac{\triangle_{)2(}}{\triangle_{)1(}} = \frac{\lambda_1 + \lambda_3 - \lambda_1\lambda_3}{\lambda_2 + \lambda_3 - \lambda_2\lambda_3}.$$

We would have found a similar result for the elementary regression coefficient (that is for the coefficients obtained by minimising in the direction of the axes of one of the variates). Now suppose that we add a variate x_4 which has nothing at all to do with the other variates. The easiest way to express this in our formulae is to say that x_4 has no random part, i. e. $\lambda_4 = 0$, and that its systematic part is uncorrelated with all the other variates. Let us first make use only of the latter property but temporarily, for the symmetry of the formulae, preserve the letter λ_4. We now have $F_{123} = F_{124} = F_{134} = F_{234} = 0$, $F_{12} = F_{13} = F_{23} = 0$ and by (8.1), since now $\varrho_{ij} = 0$ when $i \neq j$, $F_{14} = (1 - \lambda_1)(1 - \lambda_4)$, $F_{24} = (1 - \lambda_2)(1 - \lambda_4)$, $F_{34} = (1 - \lambda_3)(1 - \lambda_4)$; consequently

$$\begin{aligned}
\triangle_{)1(} = \triangle_{234} = &\lambda_2(1 - \lambda_3)(1 - \lambda_4) + \lambda_3(1 - \lambda_2)(1 - \lambda_4) \\
&+ \lambda_4(1 - \lambda_2)(1 - \lambda_3) + \lambda_2\lambda_3(1 - \lambda_4) \\
&+ \lambda_2\lambda_4(1 - \lambda_3) + \lambda_3\lambda_4(1 - \lambda_2) \\
&+ \lambda_2\lambda_3\lambda_4
\end{aligned}$$

Collecting the terms of this expression we find that it reduces to

$$(12.3) \qquad \triangle_{234} = (1 - \lambda_4) \cdot \triangle_{23} + \lambda_4.$$

Therefore if $\lambda_4 = 0$, we get $\triangle_{234} = \triangle_{23}$. And similarly for $\triangle_{)2(} = \triangle_{134}$, etc. In other words: *The disturbances will tend to give the same spurious value to $\triangle_{234}$ as to $\triangle_{23}$.* The same would apply to the scatterances in other subsets. This means that the diagonal regression is virtually unchanged by the inclusion of the extraneous variate. And the same applies to the elementary regression.

This effect appears in its purest form when the new variate is completely uncorrelated with the systematic parts of the old variates. But the *tendency* will be the same even if there should be some correlation between the new and old systematic variates.

A regression coefficient will of course also be relatively stable under the inclusion of an extraneous when the regression slope in question express a *significant* connection between the variates (when the original set was *not* multiple collinear), but

in this case the stability is less interesting from the confluency view-point.

13. THE DEGRADATION EFFECT.

When we have included so many variates in the regression analysis that a situation is reached where the systematic parts of the variates are multilinearly connected, the empirical regression coefficient will, as we have seen, depend essentially on the disturbances; they will for instance not reveal the true magnitudes which the regression coefficients have in the lower sets that were exactly and simply collinear. Can anything be said in general about what values the regression coefficients in the multilinear sets tend to assume?

I believe that some such general rules may be formulated, although I have not at the moment any exact proof of it. It seems that there exists a *degradation effect* of regression coefficients in the sense that when too many variates are included so that we get a multiply collinear set, the coefficients tend back to the gross value they had in the lowest sets with the poorest fit, that is to say the coefficients tend back to the values they had *before* the systematic influence connected with the other relevant variates was eliminated.

Suppose, for instance, that there are four observational variates, and that there exist good linear relations in all the three-sets, while there is only very little correlation in the two-seets. The elementary regression coefficient of x_1 on x_2 determined in the set (12) — we denote it b_{12} — will then, to a large extent, be determined by exterior influences. Adding the variate 3 we get a regression coefficient of x_1 on x_2, namely b_{123}, that may be entirely different from b_{12}; b_{12} may for instance indicate a positive connection between x_1 and x_2, while b_{123} may indicate a negative connection. On our assumptions the fit in the three-set (123) is good, and b_{123} would actually be a significant expression for the connection between x_1 and x_2, but b_{12} would not express any systematic connection between x_1 and x_2.

Now suppose that we add still another variate x_4. On the assumption that not only (123) but also one, or which amounts to the same, all of the other three-sets are linearly connected, the *degradation* of the coefficient considered would now take place: There would be a tendency for $b_{12.34}$ to come back to

the value b_{12}. Instead of improving the result by adding x_4 we come back to the poorest of all the results obtained so far.

What happens is very much the same as what happened to the man who climbed up a ladder, and insisted on taking still another step after he had reached the top of the ladder.

The above rule can be illustrated and its general validity rendered plausible by the following considerations. Suppose that we have 4 observational variates x_i which are made up as linear combinations of the 6 basic variates y_K in the following way

$$\begin{aligned}
x_1 &= y_1 + && 0.1\ y_3 \\
x_2 &= y_2 + && 0.1\ y_4 \\
x_3 &= y_1 + y_2 + 0.1\ y_5 \\
x_4 &= y_1 - y_2 + 0.1\ y_6
\end{aligned}$$

(13. 1)

In other words y_1 and y_2 constitute that part of the basic variates that produce the systematic connection within the observational set (1234), and y_3, y_4, y_5 and y_6 constitute disturbances.

If the variates y are determined by independent random drawings, the observed x will show certain linear connections. The systematic parts x_i' of the x_i will indeed be

$$\begin{aligned}
x_1' &= y_1 \\
x_2' &= y_2 \\
x_3' &= y_1 + y_2 \\
x_4' &= y_1 - y_2
\end{aligned}$$

(13. 2)

And eliminating the basic variates y_1 and y_2 from these equations we see that any three-set of the systematic parts of the observational variates will satisfy a linear equation. Taking the coefficients of these equations as the "true" regression coefficients we see that the "true" regression equations will be:·

(13. 3)

In the set not containing No. 1							$2x_2$	$-x_3$	$+x_4$	$=0$
"	"	"	"	"	"	2	$2x_1$		$-x_3$	$-x_4$	$=0$
"	"	"	"	"	"	3	x_1	$-x_2$		$-x_4$	$=0$
"	"	"	"	"	"	4	x_1	$+x_2$	$-x_3$		$=0$

Of these four equations it is, of course, only two that are linearly dependent.

By (7.5) the observed crossmoments will be

$$
\begin{aligned}
m_{12} &= & s_{12} + 0.1 \ (s_{14} + s_{23}) + 0.01 \cdot s_{34} \\
m_{13} &= & 1 \ + s_{12} + 0.1 \ (s_{15} + s_{13} + s_{23}) + 0.01 \ s_{35} \\
m_{23} &= & 1 \ + s_{12} + 0.1 \ (s_{25} + s_{14} + s_{24}) + 0.01 \ s_{45} \\
m_{14} &= & 1 \ - s_{12} + 0.1 \ (s_{16} + s_{13} - s_{23}) + 0.01 \ s_{36} \\
(13.4) \quad m_{24} &= -1 \ + s_{12} + 0.1 \ (s_{26} + s_{14} - s_{24}) + 0.01 \ s_{46} \\
m_{34} &= & + 0.1 \ (s_{16} + s_{26} + s_{15} - s_{25}) + 0.01 \ s_{56} \\
m_{11} &= & 1.01 + 0.2 \ s_{13} \\
m_{22} &= & 1.01 + 0.2 \ s_{24} \\
m_{33} &= & 2.01 + 2s_{12} \ + 0.2 \ (s_{15} + s_{25}) \\
m_{44} &= & 2.01 - 2s_{12} \ + 0.2 \ (s_{16} - s_{26})
\end{aligned}
$$

If we can assume that the basic variates are exactly uncorrelated, (13.4) reduces to

$$
\begin{aligned}
m_{12} &= 0 \\
m_{13} &= 1 \\
m_{23} &= 1 \\
m_{14} &= 1 \\
m_{24} &= -1 \\
m_{34} &= 0 \\
m_{11} &= 1.01 \\
m_{22} &= 1.01 \\
m_{33} &= 2.01 \\
m_{44} &= 2.01
\end{aligned}
$$

(13.5)

Determining in an actual case the values of the basic variates y by random drawings we would, of course, not get exact non correlation between them, but some small accidental intercorrelations. These intercorrelations between the basic variates would not, however, have a very great effect on the regressions computed. For the three-sets this is seen by comparing (13.6) with (13.7). (For the s were used the values obtained in the experiment which is discussed in more detail in Section 23). It is seen that (13.6) and (13.7) give essentially the same result. And this result is also very close to the "true" regression coefficients given by (13.3). In order to make the comparison with (13.3) easy the coefficients in (13.6) and (13.7) are reduced so as to give the same absolute row sums as in (13.3).

TABLE (13.6).

In set not containing the variate No.	Diagonal regression coefficients computed on the basis of the complete moments (13.4).			
	x_1	x_2	x_3	x_4
1	0	+ 1.985	— 1.021	+ 0.994
2	+ 2.011	0	— 0.991	— 0.998
3	+ 1.028	— 0.985	0	— 0.987
4	+ 1.008	+ 0.999	— 0.993	0

TABLE (13.7).

In set not containing the variate No.	Diagonal regression coefficients computed on the basis of the abbreviate moments (13.5).			
	x_1	x_2	x_3	x_4
1	0	+ 1.990	— 1.005	+ 1.005
2	+ 1.990	0	— 1.005	— 1.005
3	+ 1.002	— 1.002	0	— 0.996
4	+ 1.002	+ 1.002	— 0.996	0

When we consider the four-set (1234) we find some, although not a very great discrepancy between the result obtained by using the complete moments (13.4) and the abbreviated moments (13.5). This is seen by comparing (13.8) with (13.9).

TABLE (13.8). ADJOINT MOMENT MATRIX COMPUTED ON THE BASIS OF THE COMPLETE MOMENTS (13.4).

$\hat{m}_{ij}$	$j = 1$	2	3	4
$i = 1$	0.067	0.007	— 0.036	— 0.029
2		0.066	— 0.037	0.029
3			0.037	— 0.001
4				0.029

TABLE (13.9). ADJOINT MOMENT MATRIX COMPUTED ON THE BASIS OF THE ABBREVIATED MOMENTS (13.5).

	$j = 1$	2	3	4
$i = 1$	0.061	0.000	— 0.030	— 0.030
2		0.061	— 0.030	0.030
3			0.030	0.000
4				0.030

Of course neither (13. 8) nor (13. 9) can be compared with any "true" table of coefficients, since it has no meaning to speak of a regression equation in the four-set (1234).

If the degradation effect manifests itself in the way suggested above, we would expect that the empirical regression coefficients determined in the set (1234), in other words, in the set that is *too large* to give a meaning to the coefficients (because the set is multicollinear), should be more or less equal to the coefficients determined in the sets that are *too small* to give the correct *net* coefficients (because the sets do not include all significant variates). That this is actually so, will appear, for instance, by a comparison between the elementary regression coefficient of x_1 on x_2 in the sets (12), (124) and (1234) respectively. Using the complete moments (13. 4) we find

$$
\begin{aligned}
b_{12} &= - \ 0.120 \\
b_{12.4} &= + \ 0.919 \\
b_{12.34} &= - \ 0.112
\end{aligned}
$$
(13. 10)

The coefficient determined in the four-set is here virtually equal to the corresponding gross regression coefficients determined in the two-set. And both these regression coefficients, both that in the two-set and that in the four-set, are entirely different from the "true" coefficients existing in the three-set and being equal to $+ 1$.

In the more complete discussion of this numerical example in Sections 23 and 24 we shall see that this degradation effect is present all over.

PART III. BUNCH ANALYSIS.

14. REGRESSION STABILITY UNDER A CHANGE IN THE MINIMALISATION DIRECTION.

Let us sum up the main conclusions of the analysis of the last Sections. Each observational variate is considered as made up of a systematic part and a disturbance. The first is that part which we may hope to "explain" by considering simultaneously the variations of several observational variates, the "explanation" taking statistically the form of one or more regression equations between the variates. The other part is that which we cannot hope to "explain" in such a way. Judging the

significance of a coefficient in such a regression involves therefore two questions, one regarding the systematic variations, the other regarding disturbances: (I) Does the coefficient represent a *net* relation between the two variates which it connects? In other words, are all other systematic influences eliminated? (II) What is the precision of the determination? That is to say, how much discrepancy must we be prepared to reckon with between the actually computed value of the coefficient and that value which would have emerged if no disturbance had been present?

This distinction between systematic variations and disturbances being adopted, the purpose of any attempt at statistical determination of regression coefficients may be formulated thus: We want to compute a coefficient of which it can be said in the first place that it represents a *net* relation, and in the second place that it is very improbable that it is widely different from that result that would have emerged if no disturbance had been present.

If we have reason to believe that an empirically determined intercoefficient in a regression equation does not represent a net relation in the above sense, we may try to include in the regression equation one or more new variates and then determine the intercoefficient as it appears in the new equation. But by so doing we must however be very careful. There are at least four different things to take account of in this connection.

(1) In the first place the new variate may contain a component that is systematically connected with the other variates, it is just because of this component that the idea presents itself to include the new variate as an element in the analysis. But the presence of some such systematic component in the new variate is in itself not a sufficient ground to include this variate in the analysis.

(2) Account must also be taken of the fact that a disturbing element is always introduced as a part of the new variate. This disturbing element will lessen the precision of the total result obtained; and the loss on this account may be larger than the gain produced by the introduction of the systematic part of the new variate.

(3) The situation may further be such that not only is a new disturbing element introduced by the new variate, but the

inclusion of this variate may create a situation where the disturbing elements contained in all the variates is given a *larger opportunity of influencing the result than before.* The whole regression technique, built on moment determinants, etc. become indeed all the weaker and all the more sensitive to random disturbances the larger the number of variates included. This increase in the sensitivity to random disturbances is particularly great if in the new set of variates there are two (or more) subsets whose systematic parts are fairly well linearly connected. In this case the new set will be near to a multiply collinear set; and the regression coefficients consequently depend essentially on the random disturbances. We get a degradation effect as discussed in Section 13.

(4) The tendency towards nonsense result mentioned in (3) will be lessened so far as a certain intercoefficient is concerned, if the structural relations between the variates are of a special sort, namely such that the intercoefficient in question is nearly the same in all the subsets in that big set which comes near to being multiply collinear. This is the persistency effect studied in Section 11.

The kind of change we shall get when we include the new variate will depend on the *relative strenghts* of all these four tendencies.

Both the size and the precision of a given intercoefficient may change. The precision may for instance be so much weakened that we would have been better off by not including the new variate. It might have been better deliberately to leave some of the *systematic* bias in the coefficient in order to be better protected against the *random disturbances*. This is quite conceivable if the object of the statistical investigation is — as we formulated it above — to get a regression coefficient of which it can be said that it is very improbable that it is widely different from the "true" regression. To use an illustration. In target shooting the result depends, not only on the correct aiming but just as much on the steadiness with which one pulls the trigger. If for some particular reason it is impossible to pull the trigger steadily when one aims *exactly* at the target, it is quite conceivable that it would be better deliberately to aim a little on the side of the target. And so in statistical analysis it may be found safer deliberately to

leave some bias in the regression coefficients by not including a certain variate in the analysis.

Does there exist any empirical criterion which can tell us whether — when all these various factors are taken into account — a certain variate *ought* to be included or not?

Let us for a moment revert to the two-variates example discussed in Section 9. We found that under the assumptions (I), (II) and (III') of Section 7 the "true" regression slope must lie between the regression slope obtained by minimising in the x_1 direction and that obtained by minimising in the x_2 direction. These two slopes form *limits* between which the true slope must lie whenever the assumptions specified hold good. But there is nothing in the observed correlation matrix (here consisting only of the correlation coefficient $r_{12}^{''}$ and the unit elements in the diagonal) which permits to choose between the above two limits, or to fix any number intermediate between them. Thus it is when, and only when, there is a good agreement between results obtained by the two minimalisations, that the application of the assumptions (I), (II) and (III') permit to draw any definite conclusions about the "true" regression slopes.

Even without carrying a similar analysis through exactly in the general case of n variates there is one conclusion which we can draw immediately. If the random intensities are comparatively small and the systematic variates in the set considered are linearly connected, then there must necessarily be a small disagreement between the results obtained by minimalisation in the various directions. In other words, a good agreement in this sense is a *necessary* condition for a situation where we are allowed to conclude that the shape of the scatter reveals anything definite about the "true" linear connection between the variates. In the case of many variates this condition may not be rigorously sufficient. More precisely expressed: if several determinations of a given regression slope — for instance the one between x_1 and x_2 — is made by minimalisations in different directions in the big set, it is conceivable that although the big set is multiply collinear so that each of the results obtained is influenced primarily by the disturbing intensities (that is each result has fictitious determinateness), it may still happen that a number of them coincide more or less by pure chance. The disturbing intensities may be so distributed

that some agreement in the result is obtained. In practice such a situation will however be very improbable. It will be all the more improbable the greater the number of different determinations that coincide and the more perfect multicollinearity that exists in the big set. The regression coefficients will then almost certainly change violently with the direction of the minimalisation.

On the other hand, each determination of a given regression slope may be looked upon as answering one particular question which we have put to the data. If the data continues to give consistent answers when asked in different ways, it seems plausible to accept this as a criterion that there is something significant in these answers.

Therefore, if we study a given regression coefficient, say the one between x_1 and x_2, and include more and more variates into the analysis, we should expect that the cluster of the results obtained by minimalisation in the various directions will become tighter and tighter as new and really relevant variates are included, but we should also expect that the cluster will suddenly "explode" when some variate is introduced which makes the set multiply collinear. This is the first main idea of the confluence technique which we are now going to discuss.

An essential point in this connection is *that each regression slope is treated separately in all the subsets and for all possible minimalisation directions*. Already in Sections 5 and 6 we have made some use of the idea of regression stability under a change in the minimalisation direction, but there we condensed all the information in one single testparameter for each selection of the variates. Now we are to study the *dispersion* of the individual results. The various minimalisation directions are, so to speak, considered as elements in a "sample", and from the organisation or lack of organisation in this sample we are going to judge the significance of the result. This will prove to be a powerful tool of analysis.

The second main idea is that the *spread* of the results obtained by the minimalisation in different directions is all the way through compared with the *average value* of these results. In other words, we combine the study of net slopes with the study of precision. This combination is one of the essential features which make the present method superior to those described in

Part I, where the criteria used did not involve the slopes. The present procedure may perhaps be called *bunch analysis* on account of the particular graphical form in which it is natural to express the results for analysis. This graphical form is explained in Section 16.

It is on purpose that I have not attempted to give any formal and rigorous definition of the "probability" for a specified result obtained by the different minimalisations. Such a formal definition may indeed be obtained by starting from many *different* types of abstract schemes. Each scheme will lead to a particular definition of the probability in question. By focussing too much attention on the exact definition of the probability there is some risk that one will forget the very relative and limited meaning which must always attach to such a numerical computation of a "probability". It is indeed only in a very special meaning that any such probability can be said to measure the "significance" of the results. At least, to start with, I believe it will be a better application of time and energy to work experimentally with the method and rely on one's intuitive judgement of whether a given spread in the various determinations of a given regression coefficient is reasonable or not.

15. THE TILLING TECHNIQUE. CONCENTRIC NUMBERING.

The carrying through of the analysis whose main ideas are indicated in the last Section necessitates the computation of all the elementary regressions in all possible subsets. In practice this is done most conveniently in normal coordinates, that is to say the correlation matrix (r_{ij}) is taken as the point of departure instead of the moment matrix (m_{ij}).

The problem in thus to compute all the adjoint elements $\hat{r}_{ij(\alpha, \beta \ldots \gamma)}$ in all possible subsets. In other words, if there are n variates in the big set, we need to compute the adjoint matrix of the n rowed correlation matrix in the big set $(12 \ldots n)$, further we need the adjoints of all the n $(n-1)$-rowed matrices obtained by leaving out one variate at a time, further we need the adjoints of all the $\binom{n}{2}$ $(n-2)$-rowed matrices obtained by leaving out two variates at a time, and so on.

On the face of it this work seems to be prohibitive, but a systematic way of doing it can be found that makes it a rela-

tively simple job. The method which I have developed for this purpose is now used extensively at the Institute in Oslo. We refer to this work as "total tilling" or shorter "tilling" of the correlation matrix.

A practical tool used in tilling — and also for many other purposes — is what may be called *concentric numbering*. It is a systematic way of numbering the $\binom{n}{p}$ combinations that may be formed by selecting in all possible ways p elements out of n. For brevity the biomial coefficient $\binom{n}{p}$ is used to indicate, not only the number of such combinations, but also symbolically to denote the operation of numbering.

The concentric numbering is built up in such a way that if a new element is added, that is if we go from n to $n+1$, the list of combinations is simply *elongated*, without inserting any new numbers between those already written. This is a practical advantage in statistical work, where it will frequently be found necessary to include new variates in the course of the investigation.

The concentric numbering is defined by recurrence. First concentric numbering of the n elements $1, 2 \ldots n$, one at a time, that is for $p = 1$, is simply defined as the natural sequence of these n numbers, thus the concentric numbering $\binom{n}{1}$ consist of the n ordinals $1, 2 \ldots n$. On the other hand there is only one complex which can be formed of the n elements $1, 2 \ldots n$ taken n at a time. Thus in the concentric numbering $\binom{n}{n}$ there is only 1 ordinal, and it is written $12 \ldots n$.

These definitions of the concentric numberings $\binom{n}{1}$ and $\binom{n}{n}$ being laid down, we define concentric numbering $\binom{n}{p}$ as obtained by first writing the list of ordinals occurring in the concentric numbering $\binom{n-1}{p}$ and then elongating the list by writing the ordinals of concentric numbering $\binom{n-1}{p-1}$, adding to each of the latter ordinals the letter n. For six elements this gives for instance:

$\binom{6}{1}$	$\binom{6}{2}$	$\binom{6}{3}$	$\binom{6}{4}$	$\binom{6}{5}$	$\binom{6}{6}$
1	12	123	1234	12345	123456
2	13	124	1235	12346	
3	23	134	1245	12356	
4	14	234	1345	12456	
5	24	125	2345	13456	
6	34	135	1236	23456	
	15	235	1246		
	25	145	1346		
	35	245	2346		
	45	345	1256		
	16	126	1356		
	26	136	2356		
	36	236	1456		
	46 ·	146	2456		
	56	246	3456		
		346			
		156			
		256			
		356			
		456			

In order to carry the tilling technique through we first prepare $2^n - n - 1$ adjunction tables, namely $\binom{n}{2}$ 2-rowed tables, $\binom{n}{3}$ 3-rowed tables, etc. Each such table corresponds to a given combination in the concentric numbering of the subsets, and the cells of such a table shall receive the results of the adjunction within this subset. In other words, each such table is going to be a table of an adjoint correlation matrix. Each row and column in such a table is numbered according to the place which the elements in question hold in the big set. Thus the 4 three rowed tables in a four-variate problem will be

	1	2	3
1			
2			
3			

	1	2	4
1			
2			
4			
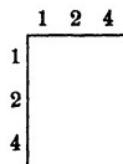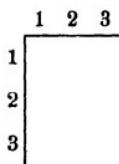
	1	3	4
1			
3			
4			

	2	3	4
2			
3			
4			

The elements in, say, the second of these tables are defined as those obtained by first forming the three-rowed correlation matrix consisting of the rows 124 and the columns 124 from the original four-rowed correlation matrix, and then taking the adjoint of this three-rowed matrix. The last element in the second row of the adjoint considered will for instance be

$$(15.1) \qquad r_{24(124)} = -(r_{42} - r_{41}\,r_{12}),$$

These tables may for brevity be called the *tilling tables* Their elements may be called the tilling elements. The tilling tables are symmetric because

$$(15.2) \qquad r_{ij(\alpha,\beta\ldots\gamma)} = r_{ji(\alpha,\beta\ldots\gamma)}.$$

As an extra bottom row in each tilling table we provide space for the numbers obtained by taking the product sum of the elements in each column of the tilling table with the corresponding elements in the original correlation table. Of course this product sum will be nothing but the value of the determinant $\triangle_{\alpha\beta\ldots\gamma}$ where $(\alpha\beta\ldots\gamma)$ is the subset considered, (i. e. the concentric number on the tilling table in question). The numbers entered in the bottom row of a given tilling table ought therefore all to be equal (apart from inaccuracies due to the fact that the computations are carried through with a limited number of decimal places). This is taken as a check on the computation; and at the same time it serves to compute the scatterance in the set $(\alpha\beta\ldots\gamma)$ to which the table refers.

The check in question can also be performed in the following way: In an extra column to the right in each tilling table we put down the sum of the elements in the various rows of the corresponding correlation matrix itself, that is, in the tilling table $(\alpha,\beta\ldots\gamma)$ we put down the sums

$$(15.3) \quad r_{\varkappa 0(\alpha,\beta\ldots\gamma)} = r_{\varkappa\alpha} + r_{\varkappa\beta} + \ldots + r_{\varkappa\gamma} \qquad (\varkappa = \alpha,\beta\ldots\gamma).$$

Then we take the product sum of each column in this tilling table with the column consisting of the numbers $r_{\alpha o(\alpha, \beta \ldots \gamma)}$, $r_{\beta o(\alpha, \beta \ldots \gamma)} \cdots r_{\gamma o(\alpha \beta \ldots \gamma)}$. This product sum ought also to be equal to $\triangle_{\alpha\beta \ldots \gamma}$. And this applies no matter which one of the columns of the tilling table we are multiplying with. Indeed, by the product summation in question the result arising from any term in the right member of (15.3) will be zero, except the result from one special of the terms, namely the one corresponding to the tilling table column with which we are multiplying; and this latter result will be $\triangle_{\alpha\beta \ldots \gamma}$. This check is very reliable because it really amounts to testing each column in the tilling table by *all* the columns in the original matrix. But it involves a little extra work, namely the formation of the sums (15.3). In most cases we only use the check based on the elementary columns of the given correlation matrix.

Since the tilling tables are symmetric, only the diagonal and one of the two triangles are filled in; as a rule we use the north-east triangle. A "row" or a "column" must then be interpreted as a broken line reflected under 45° on the diagonal.

The elements in the tilling tables may be built up systematically starting with the two-rowed tables, by means of these one computes the three-rowed tables, etc. The technique is as follows.

When the tilling tables of a certain level (for instance the three-rowed tables) are computed and checked, all the diagonal elements in the tables of the next higher level are first filled in. All these are scatterances (principal minors of the original correlation matrix), and are therefore already computed. They are found in the bottom rows of the tables of the lower level. The problem is therefore only to compute elements of the form $r_{ij(\alpha\beta \ldots \gamma)}$ where $i \neq j$. Let ν be the number of affixes in the set $(\alpha\beta \ldots \gamma)$. Further, let i stand as the p·th number in the sequence $(\alpha, \beta \ldots \gamma)$, and j as the q·th. For instance if $(\alpha, \beta \ldots \gamma)$ is the set (12567) hence $\nu = 5$, and if $i = 2$, $j = 5$, we have $p = 2$ $q = 3$. The numbers p and q being thus defined, the element $r_{ij(\alpha, \beta \ldots \gamma)}$ is equal to $(-)^{p+q}$ times the determinant:

$$
\begin{array}{c|}
 & \overline{\alpha \quad \beta \ldots)j(\ldots \gamma} \\
\alpha & \\
\beta & \\
\vdots & \\
)i(& \\
\vdots & \\
\gamma & \\
\end{array}
$$

(15. 4)

The symbol (15. 4) denotes the determinant formed by the $(\nu-1)$ rows $\alpha, \beta \ldots)i(\ldots \gamma$ and the $(\nu-1)$ columns $\alpha, \beta \ldots)j(\ldots \gamma$ of the original correlation determinant; the inverted parenthesis) (meaning "exclusion of".

Since $i \neq j$, we know that the row j and the column i actually occur in (15. 4). If $i < j$ we may therefore write (15. 4) more explicitly in the form

$$
\begin{array}{c|}
 & \overline{\alpha \quad \beta \ldots i \ldots)j(\ldots \gamma} \\
\alpha & \\
\beta & \\
\vdots & \\
)i(& \\
\vdots & \\
j & \\
\vdots & \\
\gamma & \\
\end{array}
$$

(15. 5)

Let us develop the $(\nu-1)$-rowed determinant (15. 5) according to the row j. Since j was the q-th number in the sequence $\alpha, \beta \ldots \gamma$, it will be the $(q-1)$th number in the sequence $\alpha, \beta \ldots)i(\ldots \gamma$ (when $i < j$). The first term of the expansion of (15. 5) will therefore be $r_{\alpha j}$ times $(-)^{(q-1)+1}$ times the determinant

$$
\begin{array}{c|}
 & \overline{)\alpha(\quad \beta \ldots i \ldots)j(\ldots \gamma} \\
\alpha & \\
\beta & \\
\vdots & \\
)i(& \\
\vdots & \\
)j(& \\
\vdots & \\
\gamma & \\
\end{array}
$$

(15. 6)

The next term will be $r_{\beta j}$ times $(-)^{(q-1)+2}$ times the determinant obtained from (15. 6) when in the head line $)\alpha(\beta\ldots$ is replaced by $\alpha)\beta(\ldots$. And so on. Apart from the sign the various determinants occurring in this *development is to be found in one of the tilling tables of the next lower level*. The determinant (15. 6) for instance is nothing but

(15. 7) $(-)^{p+1}\hat{r}_{i\alpha(\alpha\beta\ldots)j(\ldots\gamma)}$ (when $i < j$)

(15. 5) is consequently equal to

(15. 8) $(-)^{p+q+1}\sum\limits_{\varkappa=\alpha,\beta\ldots)j(\ldots\gamma}\hat{r}_{i\varkappa(\alpha,\beta\ldots)j(\ldots\gamma)}\cdot r_{\varkappa j}$

and therefore

(15. 9) $\hat{r}_{ij(\alpha,\beta\ldots\gamma)}=-\sum\limits_{\varkappa=\alpha,\beta\ldots)j(\ldots\gamma}\hat{r}_{i\varkappa(\alpha,\beta\ldots)j(\ldots\gamma)}r_{\varkappa j}$ $(i\neq j)$

This formula was developed on the assumption that $i < j$, but it obviously holds good also if $i > j$, the only difference in this latter case is that the sign factor of (15. 6) will now be $(-)^{q+1}$ instead of $(-)^q$, and that of (15. 7) will be $(-)^p$ instead of $(-)^{p+1}$, so that the sign factor of (15. 8) and hence of (15. 9) is unchanged. Since all the tilling tables are symmetric by virtue of their definition we see that we may interchange i and j in the right member of (15. 9).

The formula (15. 9) is capable of a very mechanical and easy application. As an example, suppose that it is wanted to compute $\hat{r}_{25(12567)}$. In the tilling table (12567) the row 2 and the column 5 are covered with cardboard or metal strips. This leaves the four figures 1567 to be read in the left margin. We look up the already computed tilling table for this set (1567). Here we consider the column 5 (i. e. that column which was *covered* in the table (12567)). We take the product sum of this column with the column 2 in the original matrix, the result — with the sign changed — is the element sought, namely $\hat{r}_{25(12567)}$.

For this work it is convenient to keep each of the columns in the original correlation matrix written (or better typewritten) on a separate strip of paper or cardboard. The pairing of the columns which are to be product-summed is then an easy matter. It is a particular advantage that there is no fuss with the *signs* in the formula (15. 9). One simply takes the

product sum of numbers that are already written in the tilling tables or in the original correlation matrix, and then *always* change the sign of the final result.

When all the elements of the tables of the ν-rowed level are thus computed, the tables are checked as explained above and one proceeds to the tables of the next higher level, etc. In Section 23 is given a complete example of these computations.

It will be noticed that all the tilling work is perfectly mechanised. It can indeed be done practically without any thinking. This is a big element in eliminating errors. Further, it will be noticed that in the course of the work not a single number is written that is not itself a finished result. And all the numbers are written immediately in just the place where they belong when the result is to be tabulated in a form convenient for systematic analysis of the various possible regression equations.

Our experience is that a complete tilling takes less time than the preparatory work of computing moments and correlation coefficients, which are indispensable for linear regression analysis according to any method. Once time and money have been invested in working out these basic parameters it is well worth taking the comparatively little extra trouble needed to make a complete tilling.

It is strongly to be recommended to do the complete tilling *at once*. No attempt should be made to pick out a few sets which the investigator for one reason or another believes are the most important. The essential point in the present method is just that all the sets are discussed without any preconceived ideas. All my practical experience with the method indicates that the complete tilling frequently brings out things which were not suspected at the outset. Furthermore, as the comparison of the various sets goes on, one will frequently want to skip back and forth looking now at one set, now at another. In the course of such an analysis it is a nuisance to be stopped because some of the sets are not computed. The work involved in the computation will of course also run much quicker and smoother when all of it is done systematically at once, instead of being done piecemeal.

As an example of the smoothness and steadiness with which the tilling work goes on I may mention that if six decimal places are carried, young assistants at the Institute will, when they

have become familiar with the method usually make something between 3 and 7 wrong multiplications in the course of all the tilling work in a six variate problem (which, as will be seen from (15.13) involves 1386 multiplications. And, of course, such mistakes are so immediately localised by the above mentioned checks that hardly any time is lost correcting them. The total tilling for a six variate problem carrying six decimal places usually takes about 13 hours, all checks and corrections included.

For any number of variates the time needed can be estimated on the basis of the number of multiplications involved, and this estimate will as a rule be a very close one because the work consist nearly exclusively of mechanical product summations.

The number of multiplications is as follows: There are $\binom{n}{k}$ k-rowed sets. Each of these contain $\binom{k}{2}$ elements that need to be computed, and each such computation involves $(k-1)$ multiplications. The total number of multiplications necessary in order to compute the elements are consequently

$$(15.10) \qquad \sum_{k=2}^{n} \binom{n}{k} \binom{k}{2} (k-1) = n \binom{n}{2} 2^{n-3}$$

The checks on a k-rowed table involves k^2 multiplication which gives a total of

$$(15.11) \qquad \sum_{k=2}^{n} \binom{n}{k} k^2 = n(n+1) 2^{n-2} - n$$

so that the grand total will be

$$(15.12) \qquad n \left[\binom{n}{2} 2^{n-3} + (n+1) 2^{n-2} - 1 \right]$$

For $n = 6$ we get for instance 720 multiplications for the direct computations and 666 for the checks giving a grand total of 1386. On the average we reckon about 100 multiplications of six decimal places per hour — checks, corrections and occasional rests included. The number 1386 therefore checks exactly with the above mentioned experience of the six-variate problem. The values of (15.12) for $n = 1, 2 \ldots$ are given in Table (15.13)

TABLE (15. 13). NUMBER OF MULTIPLICATIONS INVOLVED IN COMPLETE TILLING (ALL CHECKS INCLUDED).

Number of variates n	Number of multiplications according to (15. 12).
2	5
3	30
4	124
5	435
6	1 386
7	4 137
8	11 768
9	32 247
10	85 750
11	222 453
12	565 236

16. THE BUNCH MAP. SECTIONAL AND COMPLETE BUNCH MAPS.

When the tilling is done, the result should be exhibited graphically, otherwise it is difficult to get an ordered impression of the mass of information which is made available.

Consider a given ν-dimensional subset $(\alpha\beta\ldots\gamma)$, and let $i < j$ be two affixes in this set. If a regression equation is assumed to exist in the set, and if the equation is ordered in such a way that the variate No $\cdot i$ is expressed in terms of the variate No $\cdot j$, the equation will be of the form

$$(16. 1) \qquad \xi_i = B_{ij(\alpha\beta\ldots\gamma)}\xi_j + \cdots$$

where ξ_i and ξ_j are the variates (assumed normalised) and $B_{ij(\alpha\ldots\gamma)}$ a constant. This constant will assume different values according to which particular regression method we use. We consider specially the ν values obtained for B by taking successively the ν elementary regressions. The material for determining any such coefficient is present in the tilling tables; indeed, the coefficient B determined by the k-th elementary regression (the regression obtained by minimising in the direction of the ξ_k axis) is simply equal to

$$(16. 2) \qquad B_{ij(\alpha\beta\ldots\gamma)}^{(k)} = -\frac{\hat{r}_{kj(\alpha\beta\ldots\gamma)}}{\hat{r}_{ki(\alpha\beta\ldots\gamma)}}$$
$$(k = \alpha, \beta\ldots\gamma)$$

The numerator and denominator of (16. 2) are to be found in the tilling tables.

For a moment let us disregard the supplementary terms in (16.1) and think only of the connection between ξ_i and ξ_j. This connection may be represented by a straight line through origin in (ξ_i, ξ_j) coordinates as indicated in Figure 8.

For such a regression slope we get by (16.2) ν different determinations namely for $k = \alpha, \beta \ldots \gamma$. It is therefore a natural idea to draw on one and the same chart all these ν slopes, and see if they coincide fairly well.

It is however not only the slopes that interest us; it may be useful to note also the absolute size of the numerator and

Fig. 8.

denominator in the fraction (16. 2) which defines the slope. We shall later be concerned with various conclusions drawn on the basis of the absolute size of the numerator and denominator. We therefore plot the point M whose abscissa G and ordinate H are determined by the denominator and numerator respectively in (16. 2).

It will help further to give a clear picture of the situation if conventionally all the lines indicating the slopes are drawn from origin to one and the same side, say to the *right* (upwards or downwards as the case may be). This is also a simplification in the plotting work. The rule for the plotting will then simply be: Move towards the *right* on the horizontal axis a distance corresponding to the absolute value (regardless of sign) of the tilling elements that represents the denominator in (16. 2), then move *downwards* if the two elements to the right in (16. 2) have the same sign and *upwards* if they have opposite signs, the distance to be moved vertically being equal to the absolute value of the tilling element in the numerator in (16. 2). By convention we shall always let the variate $No \cdot i$ be measured along the vertical axis and the variate $No \cdot j$ in the horizontal axis when $i < j$.

For each set $(\alpha, \beta \ldots \gamma)$, and for each pair of the affixes (ij)
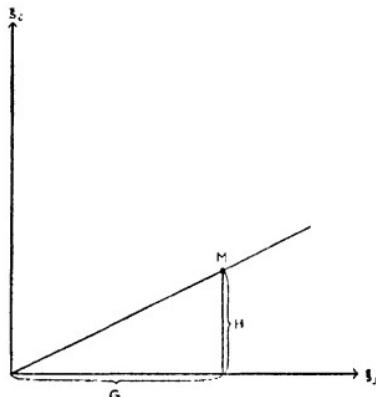
in this set a *bunch* of slope lines may thus be constructed. If the set $(\alpha, \beta \ldots \gamma)$ consists of ν variates, the bunch will contain ν beams, one for each variate. Each beam represents the result obtained by minimising in the direction of that particular variate. In the bunch representing the intercoefficient between the variates $No \cdot i$ and $No \cdot j$, the two beams $Nos \cdot i$ and j are of particular interest. They will be called the *leading beams* of the bunch.

The chart exhibiting all possible bunches in the big set $(12 \ldots n)$ we call the *bunch-map* or more explicitly the *complete* bunch map for the set $(12 \ldots n)$. The most convenient way to arrange it, if economy of space is not important, is to let all the individual bunches for a given set of variates be collected in a row, the picture of the various bunches being displayed as cells in this row.

If we do not find it necessary to investigate the slopes in all possible pairs of two variates, but are only interested in one particular such slope, for instance, the one between the variates $Nos \cdot 1$ and 2, we may limit the map to these particular bunches. Such a map we may perhaps call a *sectional* bunch map. For such a map it is practical to use a somewhat different arrangement of the cells. Figures 13 and 14 of Section 26 is an example in 7 variates.

17. THE TESTING OF A GIVEN INTERCOEFFICIENT. USEFUL, SUPERFLUOUS AND DETRIMENTAL VARIATES. THE STAR MAP.

When a new variate is tentatively added to a previously considered set, there are three fundamental possibilities to be considered. The variate may be *useful, superfluous* or *detrimental* for the purpose of the analysis as it was formulated in the beginning of Section 14. This classification may be applied either with regard to the effect of a certain variate on a given intercoefficient, for instance the regression coefficient between the variates Nos. 1 and 2, or it may be applied with regard to the effect of the variate in question on the regression equation as a whole. In the present Section we shall discuss the notions useful, superfluous and detrimental from the view-point of a given intercoefficient. In the next Section we shall consider a given regression equation as a whole.

A systematic study of the bunch map, and particularly of the change that takes place in a given bunch when we pass

from one set of variates to a more inclusive set, will furnish criteria that go a long way towards determining whether the various variates are useful, superfluous or detrimental.

We first formulate the convention that the precision of a given intercoefficient will be measured roughly by the *sprawling* of the various beams in the bunch that represent the intercoefficient in question. We may imagine that we construct the smallest sector that contains *the most important* of these beams. It is not always certain that all the beams of the bunch ought to be included in the sector measuring the precision; in certain cases one or more of the beams may by their exceptional behaviour indicate facts which make it plausible not to let them influence the sector in question. The sector would then be determined by the general behaviour of the other beams. The two leading beams in a bunch must of course always be included in that set of beams which define the precision sector.

This graphical definition of the precision of a bunch being adopted, one proceeds to the analysis of the bunch map, first going through all the cells representing the (12) coefficient, then all the cells representing the (13) coefficients, etc. Under this survey each bunch is compared with the corresponding bunch in the first subsets of the set considered. This gives rise to the following characterisation of the variate added.

A. A useful variate. If the bunch is *tightened* by the inclusion of the new variate and if the beam representing the new variate falls *inside* of the sector of the other beams, and further if the general direction of the bunch is *changed*, we conclude that the variate added is decidedly relevant. There is no doubt that it must be considered as useful for the determination of the slope in question. The behaviour of the (12) coefficient under the passage from the set (12) to (124) in Figure 11 of Section 24 has this property.

The variate would still have to be considered useful, even if the general slope did not change, but if the tightness was sharply improved and the beam of the new variate falls inside as for instance in the case of the (12) coefficient going from (12) to (123). (Fig. 11. Section 24). Both in this case and the preceding one has the new variate contributed essentially to the improvement of the fit. The only difference is that the net regres-

sion in the latter case happened to turn out about the same as the gross slope.

Even if the beam of the new variate should fall more or less outside the sector containing the other beams, the new variate would have to be considered as useful provided the bunch in general is definitely tightened, or its slope definitely changed. The examples later to be discussed will just show that if the inclusion of the new variate does change the slope markedly, there is a chance that the new beam will fall outside, particularly if it is not a long beam.

The new variate may be useful although the bunch becomes somewhat more open, namely if the general slope of the sector changes so definitely that, even taking account of the poorer precision, one gets a clear impression that the new slope is significantly different from the old. For instance, if it is a

Regression slope between variate nos. 1 and 2.

IN THE SET (I)

IN THE SET (II) OBTAINED FROM (I) BY INCLUSION OF A NEW VARIATE.
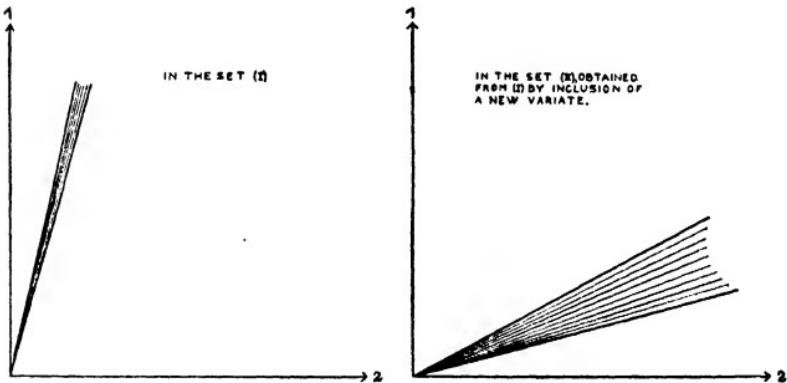
Fig. 9.

question of determining the intercoefficient between the variates Nos. 1 and 2, and we have a situation as in Figure 9, there can be no doubt that the set (II) must be preferred to the set (I)

If we have a situation as exhibited in Figure 9, we can say that it is regrettable that we need to consider a set (II) where the precision is so much poorer, as in (I), but if we are looking for the *net* intercoefficient between the variates Nos. 1 and 2, we have to run the risk that is connected with using the unprecise result in the set (II). This is just an example of a

situation where the scatterances or regression spreads or the line coefficients are not conclusive. For instance, if a number of the bunches for the other intercoefficients in the regression equation behave in a similar way, the regression spread and the line coefficient would *increase* when we go from the set (I) to the set (II).

B. A superfluous variate. The criteria of a superfluous variate are: (1) The bunch does not tighten by the inclusion of the new variate, (2) the general slope of the bunch does not change distinctly (or more specifically each of the beams in the bunch remain unchanged), (3) the beam of the new variate falls outside the sector of the other beams, (4) the beam of the new variate is much shorter than the other beams in the new bunch, (5) the beams of the other variates are not appreciably shortened by the inclusion of the new variate.

If all these criteria are simultaneously fulfilled, the variate in question is decidedly superfluous. The variate must however be considered as superfluous even if only some of the above criteria are fulfilled, particular importance must then be attached to the criteria (1) — (3).

As a rule there will in practice be a fairly good agreement between the above criteria, the only important exception being that by pure chance the beam of a superfluous variate may fall inside the general sector of the other variates.

If a variate is found superfluous by the above criteria, the conclusion should be checked by considering also the *zero slope* criterion. By this I mean the following: Let (ij) be the bunch considered (that is we considered the regression slope between the variates Nos. i and j). Let $(\alpha\beta \ldots \gamma)$ be the old set and $(\alpha\beta \ldots \gamma k)$ the new, that is, the variate added is No. k. To check the conclusion of the superfluity of $No \cdot k$ we look at the general slope of the bunches (ik) and (jk) (and possible other intercoefficients involving k) in the set $(\alpha\beta \ldots \gamma k)$. If these are degenerate, that is to say near to zero when k is the horizontal axis, or near infinity if k is the vertical axis, then this is an additional indication of the superfluity of the variate No. k. A still further confirmation of this conclusion will it be if the k-beam in the (ik) and (jk) bunches (and possibly in other bunches in the set $(\alpha\beta \ldots \gamma k)$) fall outside the sector of the other beams.

The reasons for the criteria (1) — (3) are obvious, and the criteria (4) and (5) regarding the *lenghts* of the beams are

derived by the following considerations. In any given element-
ary regression equation (in normalised variates) one of the
coefficients — namely the coefficient of that variate in whose
direction the minimalisation is done — will be a scatterance
(when the equation is written in the homogeneous form and its
coefficients determined by the elements of the adjoint correla-
tion matrix). The size of this regression coefficient will
determine the general level of the other coefficients in the
equation. A given beam in a bunch in the slope map will,
therefore, in general, be all the shorter the more perfectly the
rest of the variates in the set is linearly dependent. A glance
at a given bunch will consequently immediately give an
impression of the relative importance (for the intercoefficient
considered) of the several variates in the set. *They can be ar-
ranged in a descending order of magnitude according to the lenghts
of the beams in the bunch.* Those variates which have the longest
beams are the more important. The fact that a variate has a
beam that is very short as compared with the other beams in
the set is therefore an additional criterion of the superfluity
of this variate. (As an example see the bunch (123456) in
Figure 1 of Section 26 and the comments attached to this
bunch).

Further, the shortening of a beam as we pass from a given
set to a more inclusive one will be all the sharper the more
perfectly the *rest* of the variates in the *new* set are linearly
connected, and the poorer collinearity there is in the rest of the
variates in the *old* set.

This consideration of the *lengths* of the beams is essentially
an analysis of the same type as the study of scatterances and
minimal roots. Both analyses involve about the same sort of
information. The length of the beams is, however, only *one*
aspect of the bunch analysis. Indeed, here we combine the
study of the length of the beams with a study of their direc-
tions and with a study of the tightness of the bunch.

C. A detrimental variate. Finally, if the bunch *explodes* by the
inclusion of the new variate, that is to say, if it becomes much
less tight than before, the new variate must be considered as
detrimental. As typical examples we may consider the be-
haviour of the (12), (13) and (23) bunches when we pass from
the set (123) to (1234) in Figure 1 of Section 24.

The new variate must be considered as detrimental even

though there is, strictly speaking, no explosion, but only such a distinct decrease in the tightness and so little change in the general slope of the bunch that we cannot say definitely whether the new slope is different from the old or not.

When one gets accustomed to working with the bunch map so that one really understands "the language it speaks", one will discover that it is a veritable gold mine of information and a most powerful tool of analysis.

This classification of the variates as useful, superfluous and detrimental may be done for each intercoefficient and for the passage from any variate sets to its supersets. In order to keep track of the large number of cases that thus arise it is convenient to condense the conclusions in some sort of graphical representation. I have found the "star" map exhibited in the example of Section 24 very convenient. Here an asterisk indicates a useful variate, an empty circle indicates a superfluous variate and a blackball a detrimental variate. For instance, the asterisk on the third line in the horizontal section (234) and in the column (24) indicates that when the purpose is to determine the net intercoefficient between the variates Nos. 2 and 4 then it is correct to add the variate No. 3 to the set (24).

The above discussion refers to a comparison between a given set and its supersets, or between a given set and its subsets. It is however also of interest to compare sets on the same dimensionality level. Such a comparison may be made with regard to the behavior of the bunch of any given intercoefficient. A few examples will illustrate the situations which may here arise.

Suppose that we consider the bunch of the (12) intercoefficient, this bunch being very tight in the set (123) and also very tight in the set (124). Further suppose that the general slope of the (12) connection is markedly different in these two sets. This would be a strong indication that there exists good linear relations in both the three-sets considered, and further that this is *not the same* relation. In other words, none of the two measurements are biassed by our failure to take account of some variate. The difference is simply due to the fact that we are measuring two different things when we determine the general slope of the (12) bunch in the set (123) and in the set (124). In this case I shall say that the change in the general

slope of the (12) bunch as we pass from one to the other of the two sets (123) and (124) is a *multilinear effect*. Obviously in this case the big set that contains both the three-sets considered, namely (1234), is multiply collinear, and a regression equation in (1234) would therefore have no meaning.

On the other hand it may happen that the change in slope as we pass from (123) to (124) is due to the fact that both these slopes are biassed, the one in (123) because 4 is not taken account of, and the one in (124) because 3 is not taken account of. In this case I shall say that the change in slope is a *gross slope effect*. This is just the case where the correct solution is to unite the sets to a bigger set, namely (1234) and consider the regression here.

Which one of the two alternatives we have will be expressed by the behavior of the (12) bunch in the big set. If it explodes, we may take it as a sign that we have a multilinear effect, but if it tightens still more, we may conclude that we have a gross slope effect.

There are also other cases. Suppose for instance that there exists a good structural relation in (12345) and another in (12346). It would then have a meaning to consider a regression in the set (12345), and it would also have a meaning to consider one in the set (12346), but it would have no meaning to consider one in (123456). Suppose now that we consider statistically the two sets (123) and (124) and find that the (12) slope changes markedly as we pass from one to the other of these three-sets. This is of course in a sense a gross slope effect since in one case we have failed to take account of 3 and in the other we have failed to take account of 4, and both these variates are necessary to obtain a good regression equation. But which "true" coefficient is it of which the two observed results can be said to represent biassed measurements? In point of principle the question has no definite answer, because we may conceive of it either as the (12) slope in (12345) or in (12346). Both (123) and (124) are indeed contained in both these five-sets. In actual fact the (12) coefficient in (123) may be nearest to the "true" (12) coefficient in (12345) while that in (124) may be nearest to the "true" coefficient in (12346), or the (12) coefficient observed in both three-sets may be nearest to the "true" coefficient in (12345) etc. Which alternative we shall have, will depend on the amount of intercorrelation

that happened to be present between the various variates in the data at hand. We may refer to this as the *mixed case.*

If we got the idea of introducing also 5 and 6 amongst the variates tested, we would by the bunch technique, in all probability be guided towards the right conclusion, namely that (12345) and (12346) form two independent collinear sets.

18. THE TESTING OF THE VARIATE-SETS. CLOSED SETS AND ADMISSIBLE REGRESSIONS.

The star-map is in a sense the "difference" map of the bunch map. It expresses what happens each time we add a variate, while the bunch map expresses the situation that exists after the variate is added. By combining the criteria contained in these two maps we shall now discuss the significance of the various regressions — each regression taken as a whole — and thereby try to get an idea of the confluence hierarcy that exists in the variates. To a large extent this will consist in checking wether there is agreement between the conclusions reached by studying the individual intercoefficients.

It is not necessary — nor indeed possible — here to give a complete account of all possible cases that may arise. It will be sufficient to indicate the most important cases and the principles by which they are classified. One who wants to apply the method, will then no doubt himself be able to work out the detailed interpretation of the cases which he encounters.

I. A closed set. If we have a set, as, for instance, (234) in the example of Sections 23 and 24, where the star-map displays only asterisks, and where the slope map indicates a high degree of tightness all the way through, then we may conclude that all the variates belong in the set, and when these variates are taken together, they give a good fit to a linear relationship. All the coefficients in the regression equation in this set may be considered significant. Such a set will be called *closed* and the regression equation in this set an *admissible* regression equation. If there can be found two (or more sets) that appear as closed according to this criterion, and if the bigger set obtained by uniting these two sets show *explosions* in the various coefficients, in other words, if this bigger set has the features mentioned under IV below, then we may consider this as a check on the conclusion that the lower sets considered were closed.

II. A promising but not complete set. If we have a set where the star map shows asterisks throughout but where the bunch map does not yet indicate sufficient tightness in the various bunches, we conclude that all the variates included so far are significant, and should stay in, but we should continue to search for some new variate that may be able to improve the fit.

III. A promising set but with some extraneous variate(-s). If there are some circles in the set, and particularly if several of these consist in pointing out the *same* variate as superfluous, and still more if this indication is checked by the additional zero slope criterium mentioned in Section 17, then the variate (or variates) in question may be left out. Some caution ought however to be taken, because it is conceivable that, by including some entirely new variate (or variates) not yet considered, the situation may be changed in such a way that the variate that first appeared as superfluous now proves to be somewhat useful. To be more precise: The inclusion of the entirely new variate (or variates) may so clarify the situation that it becomes possible to look for *finer traits* of the regression equations. And these finer traits may indicate that the variate that was first "circled" exerts, after all, some influence. We may then interpret the circle first put on this variate to mean that it was superfluous *as compared with* one (or more) other variates which it was *more* important to take into account.

IV. An inadmissible set with closed subsets. If we have a set. as for instance (1234) in the example of Section 24, which in the star map is represented exclusively by blackballs, the set must be interpreted as multiply collinear. This conclusion is supported if we not only have a star map that indicates blackballs, but if the tightness in the total set — as shown by the bunch map — is actually very poor. This is the case in the example considered. The beams in the various bunches in the set (1234) are indeed sprawling excessively. And the conclusion is still more definite if the bunches in the subsets were actually quite tight so that we have clear cut *explosions* for all the bunches as we pass from the subsets to the total set. This is actually so in the example considered.

Of course, if we find a set that shows such a definite blackball picture, we need not interpret this as a deplorable result. If the tightness of the bunches in the subsets of the blackball-set are fairly good, the blackball-situation found may be

interpreted as indicating that the investigation is now carried to completion. It means that the law of variation in the subsets furnishes already all the information that can be got out of the observed product moments. In other words the cycles were already closed before we put all the variates together into one total set.

V. An inadmissible set with unfinished subsets. If the tightness in the bunches of the subsets is not satisfactory, the blackball-situation found in the total set indicates that the thing needed in order to obtain a better fit is not to take account of the fact that all the variates considered may change simultaneously and independently, *but to look for some entirely new variate.* And this new variate must again in turn be judged by the asterisk, circle and blackball criterium.

The application of these principles will be illustrated by the examples in Sections 23—29.

19. MEAN VALUES AND BOUNDARY VALUES OF REGRESSION COEF-FICIENTS. THE SIGNIFICANCE FACTOR Θ.

When we have settled the question of which one (or ones) of the regression equations are to be considered as admissible, we want to indicate limits of significance for the regression coefficients within each such equation. On the basis of the discussion in Sections 14 and 17 we adopt heuristically the two opposite elementary regression slopes as probable *boundaries* for the regression coefficient in question. This means that if the regression equation is written in the form (16.1) we adopt as boundaries for $B_{ij(\alpha\beta\ldots\gamma)}$ the numbers

$$(19.1) \qquad -\frac{\hat{r}_{ij(\alpha\beta\ldots\gamma)}}{\hat{r}_{ii(\alpha\beta\ldots\gamma)}} \text{ and } -\frac{\hat{r}_{jj(\alpha\beta\ldots\gamma)}}{\hat{r}_{ji(\alpha\beta\ldots\gamma)}}.$$

These boundaries are invariant for a permutation of the two variates. This is seen simply by noticing that if we interchange i and j in (19.1), each limit becomes equal to the reciprocal of what the other limit was originally. That is to say if we write the regression equation in the form

$$(19.2) \qquad \xi_j = B_{ji(\alpha\beta\ldots\gamma)} \cdot \xi_i + \ldots,$$

the boundaries for $B_{ji(\alpha\beta\ldots\gamma)}$ will be the reciprocals of the boundaries for $B_{ij(\alpha\beta\ldots\gamma)}$.

This suggests that it is more natural to indicate the significance of the regression coefficients by a *factor* of uncertainty than by an additive term (as we do when we indicate in the usual way the standard error of the regression coefficients). In view of this fact, it seems natural to adopt as the *mean* regression coefficient the *geometric* average between the absolute values of the two boundaries, and then determine the sign by means of $\hat{r}_{ij(\alpha\beta\ldots\gamma)}$. This gives

$$(19.3) \qquad B^{(\text{diag})}_{ij(\alpha\beta\ldots\gamma)} = -(\text{sgn}.\ \hat{r}_{ij(\alpha\beta\ldots\gamma)}) \sqrt{\frac{\hat{r}_{jj(\alpha\beta\ldots\gamma)}}{\hat{r}_{ii(\alpha\beta\ldots\gamma)}}}$$

as the mean regression coefficient, and

$$(19.4) \qquad \Theta_{ij(\alpha\beta\ldots\gamma)} = + \sqrt{\frac{\hat{r}^2_{ij(\alpha\beta\ldots\gamma)}}{\hat{r}_{ii(\alpha\beta\ldots\gamma)}\hat{r}_{jj(\alpha\beta\ldots\gamma)}}}$$

as the factor of significance.

The coefficient (19.3) is of course nothing but the *diagonal* mean regression coefficient between the variates Nos. i and j. Furthermore, the factor Θ is nothing but the absolute value of the familiar partial correlation coefficient. The use that is here made of it, namely as the significance factor (the "multiplicative standard error") of a diagonal regression coefficient is however not usual. Since Θ is the absolute value of a correlation coefficient, it must of course lie between 0 and 1.

It should be noticed that the significance limit Θ only applies to regression coefficients in equations *that have been recognised as admissible* by the whole bunch map and star map technique; Θ should not be taken as a criterion that is sufficient in itself.

In numerical work we shall add the significance factor in the same way as one usually adds the standard errors, putting, however, the sign $\cdot/\cdot$ between the two figures instead of $\pm$, so as to indicate that it is here a question of "multiplication or division", not a question of "addition or subtraction". The dots in $\cdot/\cdot$ may mnemotechnically be interpreted as the multiplication sign, and the fractional bar as the division sign. In the numerical examples of Sections 23—29 this notation is used.

In any given admissible subset $(\alpha\beta\ldots\gamma)$ the intercoefficient between any pair of two variates is defined by (19.3). If the equation as such in this set shall have a meaning, the various intercoefficients ought to be compatible, for instance B_{32}

ought to be equal to minus B_{12} divided by B_{13}, etc. Quite generally we ought to have

$$(19, 5) \qquad B_{ij} = -B_{kj}/B_{ki}.$$

This condition is fulfilled if the intercoefficients are defined by (19.3), provided only that the *signs* of the rows of the adjoint correlation matrix are compatible. If the signs are not compatible, an inspection of the bunch map will in any case tell which sign to use. If there is such a high degree of disorganisation that the sign is not distinctly defined by the bunch map, we have a set of variates that ought *not* to have passed through the severe tests of the bunch map and the star-map analysis.

Of course I have here made no attempt to specify exactly what the probability is that the "true" regression coefficient shall fall outside the significance limits indicated by Θ. For the reasons previously indicated I prefer to rely on the purposely vague statement that it is "very improbable" that the "true" coefficient will fall outside of these limits.

20. SOME TEST-PARAMETERS THAT MAY BE USED AS SECONDARY CRITERIA.

The complete analysis by the bunch map and star map is the ultimate test by which to judge empirically the hierarchy of linear confluency. If ever there is a conflict between the conclusion indicated by the maps and by some other empirical criteria, for instance some of the test parameters discussed in Part I, the result reached by the complete map study is to be considered as the final word.

This does not prevent certain test-parameters to be of some use as secondary criteria. We have already seen that when the slope map and star map analysis have decided about the admissible regressions, the *partial* correlation coefficients may be used to indicate significance limits for the coefficients. If these limits are computed for all possible intercoefficients and tabulated for all possible subsets in tables of the same arrangements as the tilling tables, we even get an approximate idea of one feature of the bunches, namely the angle between the two leading beams in each bunch; obviously the partial correlation coefficient is an expression for this angle. This is, however, only one of the many features that are actually taken account

of in the complete bunch map and star map analysis, so that no discussion of partial correlation coefficients, even if made in this exhaustive way, can replace the map analysis. There is also the practical consideration that it takes at least as much time to compute the complete system of partical correlation coefficients (on the basis of the tilling result) as to *plot* the bunch map. In practice one will therefore hardly ever find it worth while to compute the complete system of partial correlations. The only parameter system which it is always convenient to compute completely in practice is that involved in the tilling tables.

While the partial correlation coefficients — or which in practice amounts to the same, the factors Ω defined by (194.) — express the sprawling between the two *leading* beams in a bunch, the corresponding coefficients formed by the R's of Section 5, namely

$$(20.1) \qquad \Omega_{ij(\alpha\beta\ldots\gamma)} = + \sqrt{\frac{\hat{R}_{ij}^2}{\hat{R}_{ii}\hat{R}_{jj}}}$$

is a composite expression for the sprawling of *all* the beams in the bunch considered. In doubtful cases when it is difficult to decide by the visual inspection of the slope map whether a certain variate is useful or superfluous, or whether an equation should be taken as admissible or not, it may be found worth while to compute the parameter Ω as a supplementary indication.

A coefficient that is influenced by a still larger number of bunch characteristics is the line coefficient defined by (6.1); it expresses the sprawling of all the bunches in a given set of variates.

While the *partial* correlation coefficients — or which amounts the same, the factor Θ — as suggested above, may be of some use in the study of linear confluency, the *multiple* correlation coefficient is in my opinion of no use, or rather it is very misleading and dangerous parameters. Examples of this are mentioned in Section 1.

21. COMPATIBILITY SMOOTHING OF REGRESSION COEFFICIENTS IN OVERLAPPING SUBSETS. THE METHOD OF DIAGONAL ZEROS.

Suppose that the investigation has led up to the conclusion that there is a certain set $(12\ldots n)$ which is twofold collinear, that is to say any of the n subsets obtained by leaving out one

variate at a time is a closed set possessing a significant regression equation. If this is so, the coefficients of these n regressions ought to be *compatible* in the sense that any of these equations is deducible from any two of the other equations (by eliminating between them the variate that is lacking in the equation which it is wanted to deduce). If the regression equation within each of the subsets are determined by the diagonal regression method (or by any other empirical method), the coefficients in the set of n equations so obtained may however, not have this property exactly. The problem therefore arises to smooth the coefficients in such a particular way that the property in question is assured.

A similar but more general problem arises if the investigation has led to a certain set $(12 \ldots n)$ that is threefold collinear or more. In this general case we may formulate the problem thus: There is determined empirically an N rowed and n columned matrix of regression coefficients

$$(21.1) \qquad \| A_{ki} \| = \left\| \begin{matrix} A_{11} \ldots A_{1n} \\ \cdots\cdots\cdots \\ A_{N1} \ldots A_{Nn} \end{matrix} \right\|$$

This matrix defines N equations of the form

$$(21.2) \qquad \sum_i A_{Ki} \xi_i = 0.$$

The coefficients A_{Ki} shall be smoothed in such a way that the matrix (21.1) becomes of rank $\varkappa$.

The fact that (21.1) is of rank $\varkappa$ means that $\varkappa$ of the equations (21.2) are independent, which in turn means that the scatter in $(\xi_1 \ldots \xi_n)$ has lost $\varkappa$ dimensions, in other words, the scatter is left with an unfolding capacity of $n-\varkappa$.

The case mentioned in the beginning of this Section is the case $\varkappa = 2, N = n$. I shall first indicate a rapid method of compatibility smoothing applicable to this special case, and in the next Section I shall give a method applicable to the general case.

In the case $\varkappa = 2, N = n$ we choose such a numbering of the equations that the equation No. K is the one where the variate $No \cdot K$ is lacking. In other words all the diagonal elements in (21.1) are zero.

Let us first normalise the coefficients so as to make them

comparable in size. A rapid and convenient way to do this is by means of the absolute-value norms. This means that we determine the sum of the absolute values of the elements in each row in (21.1), and then divide each element by the absolute row sum in the row to which the element belongs. We thus obtain a new matrix

$$(21.3) \qquad \| a_{Ki} \| = \begin{Vmatrix} a_{11} \ldots a_{1n} \\ \cdots\cdots\cdots, \\ a_{N1} \ldots a_{Nn} \end{Vmatrix}$$

where all the absolute row sums are equal to unity. This property is used as a check on the normalisation. For the purpose of checks on the other computations to be considered it is convenient to compute also the natural row sums of (21.3).

Consider one particular of the rows in (21.3), say $No \cdot K$. Let Nos. P and Q be two other rows. Multiplying the row P by some constant C and the row Q by a constant D and adding, we ought to get a new row, whose elements

$$(21.4) \qquad C a_{Pi} + D a_{Qi}$$

are proportional to those of the row K, namely a_{Ki}. If the matrix (21.3) was exactly of rank 2, there would in general be one definite ratio between C and D which we would have to select in order that (21.4) should be proportional to a_{Ki}. If the matrix (21.4) is not exactly of rank $\varkappa$, it will in general not be possible to choose the ratio between C and D so as to obtain exact proportionality; C and D can only be chosen so as to obtain the "best possible" fit. This means that some more or less plausible principle must be adopted for the choice of C and D. In the present case where all the diagonal elements a_{KK} are zero it seems plausible to choose C and D in such a way that the diagonal zero is maintained. This leads to

$$(21.5) \qquad C/D = - a_{QK}/a_{PK}$$

Hence the elements of the new row $No \cdot K$ must be proportional to

$$(21.6) \qquad a_{QK} a_{Pi} - a_{PK} a_{Qi}$$

In practical computation it is convenient to put the new row either equal to plus the expression (21.6) or equal to minus

this expression, the sign to be chosen so that the elements in the new row gets the same signs as the old. Of course we assume that the signs of (21.6) for varying i will be compatible with the signs a_{Ki}, otherwise the incompatibility between the equations treated would be so great that the attempt at reconciling them should be given up and the admissibility criteria for the regressions should be reconsidered.

For each given K, the row numbers P and Q in (21.6) may be chosen in $\binom{N-1}{2}$ different ways. This means that we obtain $\binom{N-1}{2}$ different rows which can be compared with the row $No \cdot K$. To make the comparison easier we reduce all these new rows to absolute row sum unity. In other words, we form

$$(21.7) \qquad a_{Ki}^{(PQ)} = \omega \left(a_{QK} a_{Pi} - a_{PK} a_{Qi} \right)$$

where ω is a constant so chosen that the absolute row sum becomes unity and the signs of (21.7) coincide with those of a_{Ki}. If the matrix (21.3) is of rank 2, we shall have exactly $a_{Ki}^{(PQ)} = a_{Ki}$ for any set (PQ). The distribution of the values of $a_{Ki}^{(PQ)}$ around a_{Ki} for all possible sets (PQ) is characteristic for the degree to which the matrix a_{Ki} falls short of having the property which it is the purpose of the smoothing to establish. All the values (21.7) should therefore be tabulated in the way indicated in the example of Section 25. In works of this kind the natural row sums of (21.3) and of the derived matrices should be used throughout for checking purposes.

Let

$$(21.8) \qquad \overline{a}_{Ki} = \frac{1}{\binom{N-1}{2}} \sum_{PQ} a_{Ki}^{(PQ)}$$

be the average of all the coefficients which are to be compared with a_{Ki}. If $\overline{a}_{Ki}$ is not exactly equal to a_{Ki} a compromise must be made; a_{Ki} must be somewhat modified on the basis of the information contained in the *other* regression equations. A plausible solution seems to be to adopt as the smoothed coefficient the simple average between a_{Ki} and $\overline{a}_{Ki}$, that is, the smoothed coefficient is put equal to

$$(21.9) \qquad a'_{Ki} = \frac{a_{Ki} + \overline{a}_{Ki}}{2}$$

If the matrix a'_{Ki} does not yet come close to being of rank 2, the whole process may be iterated. Since the matrix $\overline{a}_{Ki}$

has absolute row sums equal to unity, and a_{Ki} and $\bar{a}_{Ki}$ always have the same sign, the matrix a'_{Ki} must have absolute row sums equal to unity, it is therefore already in shape to be taken as the starting point for a new smoothing.

The example discussed in Section 25 shows that this process converges with an extraordinary rapidity. Already the second smoothing according to this method assured exact compatibility in the first 10 decimal places. The discussion of that example even seems to indicate that the simple arithmetic average chosen in the formula (21.9) represents in a sense an optimum choice. It seems to be the linear combination between a_{Ki} and $\bar{a}_{Ki}$ that will produce the most rapid convergency possible.

22. COMPATIBILITY SMOOTHING CONTINUED. A GENERAL METHOD.

The method of compatibility smoothing described in Section 21 has a natural application only in the case $\varkappa = 2$, $N = n$, where the a_{Ki} is a square matrix with zeros in the diagonal. Indeed, these zeros just served to define the parameters C and D. In the general case some other principle must be relied upon.

If the matrix a_{Ki} is of rank exactly $\varkappa$ it should be possible to express the row $No \cdot K$ as a linear combination of $\varkappa$ other rows, in other words we should have, for any i

$$(22.1) \qquad a_{Ki} = C_P a_{Pi} + C_Q a_{Qi} + \ldots + C_R a_{Ri}$$

where $P, Q \ldots R$ is a set of $\varkappa$ different numbers chosen in the sequence $1, 2 \ldots N$, and the $C's$ are coefficients independent of i.

How shall the coefficients C be selected in order to assure the best possible fit?

A natural idea seems to be simply to determine them as the regression coefficients of a_{Ki} on $a_{Pi}, a_{Qi} \ldots a_{Ri}$, i being the variable parameter that defines the various "observations". This leads to the following procedure.

Let the matrix (21.1) now be reduced by the square row norms, in other words put

$$(22.2) \qquad a_{Ki} = A_{Ki} / \sqrt{A_{Ki}^2 + \ldots + A_{Kn}^2}$$

so that

$$(22.3) \qquad a_{K1}^2 + \ldots + a_{Kn}^2 = 1 \text{ for all } K$$

Then form the row moments

$$(22.4) \qquad \mu_{HK} = \Sigma_i a_{Hi} a_{Ki},$$

(in particular by (22.3) $\mu_{KK} = 1$), and compute (most conveniently by the tilling technique) the adjoint elements in all $(x + 1)$ rowed subsets

$$(22.5) \qquad \hat{\mu}_{KH(UV\ldots W)}$$

$(UV\ldots W)$ being $(x + 1)$ numbers chosen in the sequence $1, 2 \ldots N$, and the adjunction being made in this set. With this notation the coefficient $a_{Ki}^{(PQ\ldots R)}$, which is obtained via the equations $(PQ\ldots R)$, and is to be compared with a_{Ki}, is defined by

$$(22.6) \quad \hat{\mu}_{KK(KPQ\ldots R)} \cdot a_{Ki}^{(PQ\ldots R)} = -\sum_{H=P, Q\ldots R} \hat{\mu}_{KH(KPQ\ldots R)} \cdot a_{Hi}$$

where H runs through all the x numbers $P, Q \ldots R$ (but not K), and $\hat{\mu}$ are the adjoints taken in the $(x + 1)$ rowed set $(KPQ\ldots R)$. There is no need to normalise the magnitudes $a_{Ki}^{(PQ\ldots R)}$ determined by (22.6) since by this formula they have already been fitted to a_{Ki} both in sign and order of magnitude.

For each K there will by (22.6) be obtained $\binom{N-1}{x}$ different values to compare with a_{Ki}. In the case of $x > 2$ it will be a little troublesome to tabulate all these values, as we did in the method of Section 21. If it is not particularly wanted to see each value, the tabulation may, however, now be omitted because the present method is such that the averaging may be done in the algebra of the formula. In order to do this we first make the convention that (22.5) shall be interpreted as zero whenever one, or both, the affixes (HK) lack in the set $(UV\ldots W)$. With this convention the summation in the right member of (22.6) may be interpreted to run through all the values $1, 2 \ldots N$ except K. This is a formal advantage when we want to perform an averaging over all possible sets $(PQ\ldots R)$. The average of the magnitudes (22.6) using the $\hat{\mu}_{KK(KPQ\ldots R)}$ as weights, will now be

$$(22.7) \qquad \overline{a}_{Ki} = -\sum_{H=1, 2\ldots)K(\ldots N} \hat{\mu}_{(KH)} a_{Hi} / \hat{\mu}_{(KK)}$$

where

$$(22.8) \qquad \hat{\mu}_{(KH)} = \Sigma_{PQ\ldots R} \hat{\mu}_{KH(KPQ\ldots R)}$$

the summation in (22. 8) running over combinations without repetition of the $\varkappa$ affixes $(PQ \ldots R)$. This is obviously the same as $\sum \hat{\mu}_{KH(UV \ldots W)}$ where the $(\varkappa+1)$ affixes $(UV \ldots W)$ run through those special combinations that contain K. By the above convention this is however in turn the same as if $(UV \ldots W)$ runs through all possible combinations, no matter whether K is present or not. We therefore have

$$(22.9) \qquad \hat{\mu}_{(KH)} = \sum_{UV \ldots W} \hat{\mu}_{KH(UV \ldots W)}$$

where (KH) are any two affixes (equal or unequal) in the set $1, 2 \ldots N$ and $\sum_{UV \ldots W}$ denotes a summation over combinations without repetition of the $(\varkappa+1)$ affixes $UV \ldots W$ picked in the set $1, 2 \ldots N$. Thus the N-rowed square matrix $\hat{\mu}_{(KH)}$ is simply formed by adding all the (KH) elements that occur in the $(\varkappa+1)$ rowed tilling tables for μ_{KH}. This matrix may be formed once for all and applied to all the averagings (22.7). When the matrix $\hat{\mu}_{(KH)}$ is formed, each of the magnitudes that is to be compared with a_{Ki} is formed as a linear compound, namely (22. 7) of *all* the $(N-1)$ coefficients a_{Hi} where $H \neq K$. In order to omit the term $H = K$ in the summation (22.7) and get the proper sign it will in practice be most convenient to make a complete square table of the matrix $- \hat{\mu}_{(KH)}$ (which, incidentally, is symmetric) divide by $\hat{\mu}_{(KK)}$ and replace all the diagonal elements by zeros. The quantities $\bar{a}_{Ki}$ are then simply formed by taking the product sum of a column in this modified matrix and a column in the original a matrix.

Since by the computation of the accumulated matrix (22. 9) each element in the $(\varkappa+1)$-rowed tilling tables is used once, and only once, various forms of checks may be applied. We may for instance verify that:

$$(22.10) \quad \left\{ \begin{array}{l} \text{The sum of the diagonal elements in } \hat{\mu}_{(KH)} \text{ is equal} \\ \text{to the sum of the diagonal elements in all the } (\varkappa+1) \\ \text{rowed tilling tables, and similarly for the north-east} \\ \text{triangle.} \end{array} \right.$$

Using the same argument as in Section 21, we put the smoothed value of a_{Ki} equal to

$$(22.11) \qquad a'_{Ki} = \frac{a_{Ki} + \bar{a}_{Ki}}{2}$$

If necessary the values a'_{Ki} may be taken as the starting point for a second smoothing, etc.

Instead of adopting the simple arithmetic average (22.11) *sans façon*, one may go to the following more elaborate procedure. Let us put

$$(22.12) \qquad a'_{Ki} = a_{Ki} - \lambda c_{Ki}$$

where

$$(22.13) \qquad c_{Ki} = a_{Ki} - \bar{a}_{Ki}$$

and λ is a parameter to be determined in such a way that the smoothing comes the nearest possible to making the matrix a'_{Ki} of rank $\varkappa$. Since a common factor of proportionality is of no avail for the smoothing of the matrix considered, (22.12) represents the most general form of linear smoothing that can be based on the two elements a_{Ki} and $\bar{a}_{Ki}$.

The moments μ'_{KH} of the new matrix a'_{Ki} will be

$$\mu'_{KH} = \sum_i a'_{Ki} a'_{Hi} = \sum_i a_{Ki} a_{Hi} - \lambda (a_{Ki} c_{Hi} + a_{Hi} c_{Ki}) + \lambda^2 c_{Ki} c_{Hi}.$$

Hence

$$(22.14) \qquad \mu'_{KH} = \mu_{KH} - \lambda \nu_{KH} + \lambda^2 \gamma_{KH}$$

where μ_{KH} is the moment matrix of the original coefficients, defined by (22.4), and

$$(22.15) \qquad \nu_{KH} = \sum_i a_{Ki} c_{Hi} + a_{Hi} c_{Ki}$$

$$(22.16) \qquad \gamma_{KH} = \sum_i c_{Ki} c_{Hi}$$

As a check on (22.15) and (22.16) we have

$$(22.17) \qquad \nu_{00} = 2 \sum_i a_{0i} c_{0i}$$

$$(22.18) \qquad \gamma_{00} = \sum_i c_{0i}^2$$

where a_{0i} and c_{0i} are column sums and ν_{00} and γ_{00} grand totals.

All the three matrices μ, ν, γ are obviously symmetric.

Since the elements of the new moment matrix are of the form (22.14), we are led to consider the N-rowed determinant

(22.19) $$F(\lambda) = |\mu_{KH} - \lambda \nu_{KH} + \lambda^2 \gamma_{KH}|$$

and its various principal minors. Let F_K be the $(N-1)$-rowed principal minor obtained from (22.19) by omitting the row and column No K, F_{KH} the $(N-2)$-rowed principal minor obtained by omitting the rows Nos. K and H and the columns Nos. K and H, etc. All these determinants are obviously positive definite for *any* real value of λ because they are moment determinants no matter what value we put for λ. This shows that any value of λ which makes one of the principal minors of (22.19) vanish, must also make all those principal minors vanish, which contain the first minor.

In order that the new coefficient matrix a' shall be of rank $\varkappa$, it is necessary and sufficient that the moment matrix μ' is of rank $\varkappa$ which means that the determinant (22.19) and all its principal minors down to and including the $(\varkappa + 1)$-rowed must vanish. In general it will not be possible to ensure this just by disposing of the single parameter λ, but in practice if the original coefficient matrix a was near to being of rank $\varkappa$, it may be possible to select a value of λ that will realise very closely the vanishing of the principle minors in question. Since (22.19) is positive definite for all values of λ, the shape of the function $F(\lambda)$ and its principal minors will be of the kind exhibited in Figure 10.

None of the functions F can *pass* zero; since they are positive definite for all λ, they can at most *touch* zero. And in a point where any of the F_K touches zero, F must also do so.

If all the functions $F_{UV \ldots W}$ of the order $(\varkappa + 1)$ vanish in

Fig. 10.

points that lie tightly together in a group, then we would obtain a good solution by putting λ equal to some central value in the group. In Section 25 we shall see a numerical example where these points lie extremely close.

If several such groups should exist, we are of course inter-

ested only in the one with smallest λ; we see indeed by (22.12) that the correction to be applied to a_{Ki} is all the smaller the smaller λ.
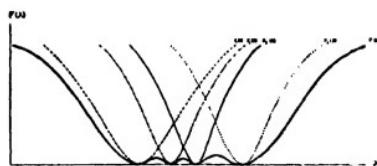
Thus from the point of view of numerical computation the problem is to plot the minors in question $F_{UV \dots W}$ over a certain range near $\lambda = 0$, and locate their zeros in this range.

In order to do this we need to develop functions of the kind (22.19) as polynomials in λ. For $N = 2$ we find

$$(22.20) \quad \begin{aligned} |\mu_{HK} - \lambda \nu_{HK} + \lambda^2 \gamma_{HK}| &= \mu_{11}\mu_{22} - \mu_{12}\mu_{21} + \\ &+ (\mu_{12}\nu_{21} + \nu_{12}\mu_{21} - \mu_{11}\nu_{22} - \nu_{11}\mu_{22})\lambda + \\ &+ (\gamma_{11}\mu_{22} + \nu_{11}\nu_{22} + \mu_{11}\gamma_{22} - \gamma_{12}\mu_{21} - \nu_{21}\nu_{12} - \mu_{12}\gamma_{21})\lambda^2 + \\ &+ (\nu_{12}\gamma_{21} + \nu_{21}\gamma_{12} - \gamma_{11}\nu_{22} - \nu_{11}\gamma_{22})\lambda^3 + \\ &+ (\gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21})\lambda^4. \end{aligned}$$

and for $N = 3$ we get

$$(22.21) \quad \begin{aligned} |\mu_{HK} - \lambda \nu_{HK} + \lambda^2 \gamma_{HK}| &= S_{\mu\mu\mu} - S_{\mu\mu\nu} \cdot \lambda + (S_{\mu\nu\nu} + S_{\mu\mu\gamma})\lambda^2 \\ &- (S_{\mu\nu\gamma} + S_{\nu\nu\nu})\lambda^3 + (S_{\mu\gamma\gamma} + S_{\nu\nu\gamma})\lambda^4 \\ &- S_{\nu\gamma\gamma}\lambda^5 + S_{\gamma\gamma\gamma}\lambda^6. \end{aligned}$$

Here $S_{\mu\nu\gamma}$ denotes the sum of all the determinants that can be formed by taking in all possible ways one column from the matrix μ, one from ν and one from γ, $S_{\mu\mu\nu}$ denotes the similar sum when two columns are taken from μ and one from ν, etc. When μ, ν and γ are written as affixes, they may be interpreted to have the *weights* 0, 1 and 2 respectively. The sum of the weights in a given term in (22.21) is equal to the exponent of λ, and the term consists of all possible S that have subscripts with total weight sum equal to the exponent of λ.

If the S in (22.21) are written out explicitly we get

$$(22.22) \quad \begin{aligned} S_{\mu\mu\mu} &= |\mu\mu\mu| \\ S_{\mu\mu\nu} &= |\mu\mu\nu| + |\mu\nu\mu| + |\nu\mu\mu| \\ S_{\mu\nu\nu} &= |\mu\nu\nu| + |\nu\mu\nu| + |\nu\nu\mu| \\ S_{\mu\mu\gamma} &= |\mu\mu\gamma| + |\mu\gamma\mu| + |\gamma\mu\mu| \\ S_{\nu\nu\nu} &= |\nu\nu\nu| \\ S_{\mu\nu\gamma} &= |\mu\nu\gamma| + |\mu\gamma\nu| + |\nu\mu\gamma| + |\nu\gamma\mu| + |\gamma\mu\nu| + |\gamma\nu\mu| \\ S_{\nu\nu\gamma} &= |\nu\nu\gamma| + |\nu\gamma\nu| + |\gamma\nu\nu| \\ S_{\mu\gamma\gamma} &= |\mu\gamma\gamma| + |\gamma\mu\gamma| + |\gamma\gamma\mu| \end{aligned}$$

Here $|\mu\mu\nu|$ denotes the determinant where the first and second columns are taken from the matrix μ, and the third column from the matric ν, and similarly for the other symbols in (22. 22). The rules indicated above are general and applies to any N.

In practice the computations are not so elaborate as they may appear from this theoretical analysis; indeed, in practice the higher terms will frequently vanish nearly exactly so that we actually need to work only with polynomials of rather low order. The example of Section 25 shows this. This example also turns out to lead nearly exactly to the simple arithmetic average which we put up heuristically in formula (22. 11).

PART IV. APPLICATIONS.

23. A CONSTRUCTED CONFLUENCY EXAMPLE IN 5 VARIATES.

To test the various procedures suggested in the preceding Sections and compare their relative merits we shall analyse a constructed numerical example and some examples drawn from actual economic data. We begin with the constructed example.

Consider four variates $x_1 \ldots x_4$ whose values in each observation point are determined by (13.1) where the y's are variates determined by random drawings. In the example each y_i observation was determined by the average of end digits in 100 consecutive drawings in the Norwegian State Lottery. All the individual observations of a given y_i as well as all the various y_i were independent in the sense that drawings that were used to determine a certain observation of a given y_i, were not used for any other purpose, that is neither for other observations of this y_i nor for observations of any of the other y_j.

The Norwegian lottery drawings are done without putting the numbers drawn back into the urn. The probability for a given end digit in the various drawings is therefore not exactly constant, but the effect is so slight that it would be without any influence on the calculations in the example.

As a fifth observational variate was introduced $x_5 = y_7$, in other words x_5 was simply itself an erratic variate determined by lottery drawings.

Since the y's in this example must be considered as "independent causes", the systematic parts x_i' of the observational

variates must be interpreted as that part of the right member in (13.1) that consist of y's occurring *at the same time* in more than one x. In other words we have (13.2), while x_5 is an observational variate quite extraneous to the whole system. Eliminating y_1 and y_2 we get the "true" regression system (13.3). We shall now apply the methods developed in the preceding Sections and see if they are capable of finding the "true" regressions (13.3) and of indicating that a regression in the big set (1234) has no meaning.

When the basic variates y are determined by lottery draw-ings, as explained above, they will, of course, not become rigorously uncorrelated. This, however, only makes our example all the more realistic. The correlation coefficients as deter-mined in a series of 100 observations of the seven variates y_i turned out to be as indicated in (23.1).

TABLE (23.1). CORRELATION COEFFICIENTS BETWEEN RANDOM VARIATES y_i

s_{ij}	$j=1$	2	3	4	5	6	7
$i=1$	1.000000	-0.132076	-0.082119	0.093466	0.109091	0.146380	-0.233315
2		1.000000	0.021295	0.019510	0.024809	0.055105	0.206633
3			1.000000	-0.139912	-0.108802	-0.069751	0.094886
4				1.000000	0.002314	0.187958	0.070017
5					1.000000	0.041966	-0.268182
6						1.000000	0.156706
7							1.000000

We shall interpret the variates y that enter in the definiton (13.1) of the x to have unit sumsquare. The moments m_{ij} of the observational variates are therefore given by (7.5).

The numerical computation of such bilinear forms is most easily carried through by first forming the matrix P_{ij} defined by

$$(23.2) \qquad P_{ij} = \sum_{k=1}^{M} p_{ik} s_{kj}$$

In other word P_{ij} is the product sum of a row in the matrix p and a column (or, since s is symmetric, a row) in s. The moment matrix m_{ij} is then calculated by

$$(23.3) \qquad m_{ij} = \sum_{k=1}^{M} p_{ik} P_{jk}.$$

In other words, m_{ij} is the product sum of a row in p a row in P.

For checking purposes it is convenient to introduce the column sums

$$(23.4) \qquad p_{oj} = \sum_{i=1}^{n} p_{ij}$$

The sums (23.4) are simply handled as an $(n+1)th$ row of the matrix p; in other words, it gives rise to the determination of an $(n+1)th$ row P_{oj} of P by the general formula (23.2). Each element in the $(n+1)th$ row of P thus determined shall at the same time be the sum of the elements in the column in which it stands, in other words, we shall have

$$(23.5) \qquad P_{oj} = \sum_{k=1}^{n} P_{kj}$$

In the computation of the moment matrix the bottom rows of p and P are handled, just as the other rows: this gives the magnitudes

$$(23.6) \qquad m_{oj} = \sum_{k=1}^{M} p_{ok} P_{jk}$$

and

$$(23.7) \qquad m_{00} = \sum_{k=1}^{M} p_{ok} P_{ok}$$

m_{oj} ought to be at the same time the sum of elements in the column j in the moment matrix, and m_{oo} the grand total of all elements.

Using this technique one easily determines directly (without computing first the individual x_i observations) the following x moments:

TABLE (23.8)   MOMENTS IN THE CONSTRUCTED EXAMPLE.

$m_{ij} = [x_i x_j]$	$j=1$	2	3	4	5
$i=1$	0.993576	− 0.121999	0.871663	1.135675	− 0.223826
2		1.013902	0.881726	− 1.117290	0.213635
3			1.772628	0.028997	− 0.053500
4				2.292407	− 0.424277
5					1.000000

And from this we get the following correlation coefficients:

TABLE (23.9). CORRELATION COEFFICIENTS IN THE
CONSTRUCTED EXAMPLE.

r_{ij}	$j = 1$	2	3	4	5
$i = 1$	1.000000	− 0.121551	0.656809	0.752502	− 0.224549
2		1.000000	0.657698	− 0.732862	0.212165
3			1.000000	0.014385	− 0.040183
4				1.000000	− 0.280223
5					1.000000

By the technique of Section 15 this gives the following tilling tables:

TABLE 1. TILLING TABLES. TWO-SETS:

$\hat{r}_{ij}$	1	2
1	1.000000	0.121551
2		1.000000
$\triangle =$	0.985225	0.985225

$\hat{r}_{ij}$	1	4
1	1.000000	− 0.752502
4		1.000000
$\triangle =$	0.433741	0.433741

$\hat{r}_{ij}$	2	5
2	1.000000	− 0.212165
5		1.000000
$\triangle =$	0.954986	0.954986

$\hat{r}_{ij}$	1	3
1	1.000000	− 0.656809
3		1.000000
$\triangle =$	0.568602	0.568602

$\hat{r}_{ij}$	2	4
2	1.000000	0.732862
4		1 000000
$\triangle =$	0.462913	0.462913

$\hat{r}_{ij}$	3	5
3	1.000000	0.040183
5		1.000000
$\triangle =$	0.998385	0.998385

$\hat{r}_{ij}$	2	3
2	1.000000	− 0.657698
3		1.000000
$\triangle =$	0.567433	0.567433

$\hat{r}_{ij}$	3	4
3	1.000000	− 0.014385
4		1.000000
$\triangle =$	0.999793	0.999793

$\hat{r}_{ij}$	4	5
4	1.000000	0.280223
5		1.000000
$\triangle =$	0.921475	0.921475

$\hat{r}_{ij}$	1	5
1	1.000000	0.224549
5		1.000000
$\triangle =$	0.949578	0.949578

TABLE 2. TILLING TABLES. THREE-SETS:

$\hat{r}_{ij}$	1	2	3
1	0.567433	0.553533	− 0.736753
2		0.568602	− 0.737534
3			0.985225
$\triangle =$	0.016245	0.016245	0.016244

$\hat{r}_{ij}$	1	3	5
1	0.998385	− 0.647786	0.198156
3		0.949578	− 0.107303
5			0.568602
$\triangle =$	0.528418	0.528418	0.528418

$\hat{r}_{ij}$	1	2	4
1	0.462913	− 0.429929	− 0.663422
2		0.433741	0.641395
4			0.985225
$\triangle =$	0.015945	0.015945	0.015945

$\hat{r}_{ij}$	2	3	5
2	0.998385	− 0.666223	− 0.238593
3		0.954986	0.179723
5			0.567433
$\triangle =$	0.509590	0.509591	0.509590

$\hat{r}_{ij}$	1	3	4
1	0.999793	− 0.645984	− 0.743054
3		0.433741	0.479865
4			0.568602
$\triangle =$	0.016355	0.016356	0.016355

$\hat{r}_{ij}$	1	4	5
1	0.921475	− 0.689578	0.013681
4		0.949578	0.111249
5			0.433741
$\triangle =$	0.399494	0.399495	0.399494

$\hat{r}_{ij}$	2	3	4
2	0.999793	− 0.668240	0.742323
3		0.462913	− 0.496387
4			0.567433
$\triangle =$	0.016273	0.016272	0.016272

$\hat{r}_{ij}$	2	4	5
2	0.921475	0.673408	− 0.006800
4		0.954986	0.124735
5			0.462913
$\triangle =$	0.426517	0.426517	0.426517

$\hat{r}_{ij}$	1	2	5
1	0.954986	0.073910	0.198760
2		0.949578	− 0.184871
5			0.985225
$\triangle =$	0.901371	0.901371	0.901370

$\hat{r}_{ij}$	3	4	5
3	0.921475	− 0.003125	0.036152
4		0.998385	0.279645
5			0.999793
$\triangle =$	0.919977	0.919977	0.919977

TABLE 3. TILLING TABLES. FOUR-SETS:

$\hat{r}_{ij}$	1	2	3	4
1	0.016273	0.001832	— 0.011739	— 0.010733
2		0.016355	— 0.012116	0.010782
3			0.015945	— 0.000275
4				0.016245
$\triangle =$	0.000263	0.000262	0.000262	0.000263

$\hat{r}_{ij}$	1	2	3	5
1	0.509590	0.505360	— 0.667867	— 0.019629
2		0.528418	— 0.680509	— 0.025978
3			0.901371	0.030631
5				0.016245
$\triangle =$	0.013910	0.013910	0.013910	0.013910

$\hat{r}_{ij}$	1	2	4	5
1	0.426517	— 0.396262	— 0.608766	0.009256
2		0.399494	0.588487	— 0.008831
4			0.901371	— 0.008970
5				0.015945
$\triangle =$	0.014507	0.014507	0.014507	0.014507

$\hat{r}_{ij}$	1	3	4	5
1	0.919977	— 0.594764	— 0.686440	— 0.009676
3		0.399494	0.443732	0.006843
4			0.528418	0.011766
5				0.016355
$\triangle =$	0.014956	0.014956	0.014957	0.014956

$\hat{r}_{ij}$	2	3	4	5
2	0.919977	— 0.616012	0.674403	— 0.030957
3		0.426517	— 0.451624	0.021279
4			0.509590	— 0.018434
5				0.016272
$\triangle =$	0.014015	0.014015	0.014015	0.014015

TABLE 4. TILLING TABLES. FIVE-SETS:

$\hat{r}_{ij}$	1	2	3	4	5
1	0.014015	0.001986	— 0.010390	— 0.009002	— 0.000214
2		0.014956	— 0.011298	0.009483	— 0.000524
3			0.014507	— 0.000531	0.000498
4				0.013910	— 0.000156
5					0.000263
△ =	0.000223	0.000223	0.000224	0.000222	0.000224

Now let us interpret the results. Consider the scatterances first. They are contained in the tilling tables. We start by considering the two-sets. None of the two-sets shows any good linearity. The best are (14) with a scatterance[1] of 0.433 and (24) with a scatterance of 0.462. If any relations worth while considering are to be found, it is clear that we must at least pass to the three-sets. Amongst the three-sets all those contained in (1234) stand out so distinctly that there can be no doubt that here is something significant. The scatterances in these sets are

	Sets	Scatterances
	123	0.0162
	124	0.0159
(23. 10)	134	0.0163
	234	0.0162

And the scatterances in the other three-sets range from about 0.4 to 0.9. Since none of the two-dimensional scatterances are small, any of the 4 three-sets in (23.10) may be accepted. This being so we ought by (V) of Section 1 to pass on to the four-sets. The four-set with the smallest scatterance is (1234); the scatterance is here 0.00026, while the other four dimensional scatterances range about 0.014. If we should let us be guided uniquely by the smallness of the scatterance, the set (1234) would seem excellent, particularly because there is a heavy drop from the subscatterances. But by (IV) of Section 1, the set (1234) must be rejected because all its subscatterances are

[1] When in the text I give abbreviated figures whose correct value are to be found in another place, I do not raise the last digit even if the first digit dropped is 5 or more. It is then easier to recognise the correct figure in the tables if it is wanted to look it up.

nearly equal, as is seen from (23. 10). The criterion indicates that (1234) may be a very dangerous set.

The four set with the second smallest scatterance is (1235). Its subscatterances are

	Sets	Scatterances
	123	0.016
(23. 11)	125	0.901
	135	0.528
	235	0.509

There is here a definite difference between the subscatterances (and all of them cannot by any means be said to be small either), so that there is no danger in accepting the set (1235). But by comparing the scatterance in (1235), namely 0.0139 with the smallest subscatterance in (1235), namely 0.016 we find that there is practically nothing gained by passing from the three-set (123) to the four set (1235). A similar analysis applies to the other four-sets containing 5.

Since the five-set (12345) contains (1234), which has already been recognised as dangerous, we cannot get any further.

The conclusions here obtained by using only the scatterances may be resumed thus: Each of the 4 three-sets contained in (1234) is collinear, making it nonsense to speak of a regression in (1234). The variate 5 is extraneous to the system. Of course, this is just the correct conclusion which we ought to obtain.

In Section 33 we shall see what perfectly absurd results are obtained by applying to the present case the usual regression technique and the significance criteria which follow from sampling theory.

24. BUNCH ANALYSIS OF THE CONSTRUCTED EXAMPLE.

Now let us apply the bunch analysis technique. This will lead to results which are still more definite and conclusive. The complete bunch map — exhibiting graphically the numbers in Table 1—4 in Section 23 — is given in Figure 11.

Let us first follow the behaviour of the bunch of the intercoefficient (12) in all possible sets. In the set (12) this bunch is very poor, which simply means that the gross correlation between the variates Nos. 1 and 2 is small. But if we add the variate No. 3, the bunch of the intercoefficient

(12) is immediately tightened in a very conspicuous way. The tightness of the (12) bunch in the original (12) set and in the set (123) cannot be compared at all. (See Figure 11). Furthermore, the 3-beam falls inside the sector of the other beams in the new set, and the 3-beam is *longer* than the other beams, and finally the other beams in the (12) bunch are shortened by the inclusion of 3. All of this points to No. 3 being essentially relevant. There can be no doubt that it is useful for the determination of the (12) intercoefficient. Adding 4 to the set (12) we get essentially a similar result: The (12) bunch is tightened, the 4 beam falls inside, the 4 beam is longer than the other beams and the other beams are shortened by the inclusion of 4. There is no doubt that 4 is useful. Thus, both 3 and 4 are essentially useful for the determination of the (12) coefficient, and when either of these variates is added, the result is a very good fit (judging the fit by the tightness of the bunch).

But the two *slopes* thus determined: the (12) slope in (123), and the (12) slope in (124) are *essentially different:* indeed, the former is negative and the latter positive. A glance at Figure 1, taking account of the tightness of both bunches and of the conspicuous difference in the slope, tells us that what is revealed by these two bunches is in all probability the (12) slope *in two different equations*. It does *not* seem possible that both these two bunches reveal the (12) slope in *one and the same equation.* This would indeed mean that the difference in slope which we have observed is only a gross-slope effect. In other words the manner in which the variates Nos. 3 and 4 has happened to vary in the material must then have been such that the empirically determined (12)-coefficient in the set (123) has been *biassed* by our not taking account of 4, and in the set (124) by our not taking account of 3. The tightness of the two observed bunches and the conspicuous difference in slopes makes it probable that no such bias exists, each of the two observed bunches representing an unbiassed coefficient, but in two different equations. This is already a rather definite indication of the multicollinearity which — from the nature of the constructed data — we know exists in the set (1234). We shall not, however, yet accept finally the conclusion of the multicollinearity in the set (1234) but continue the systematic scrutiny of the bunch map.

FIG.11

BUNCH MAP

CONSTRUCTED EXAMPLE

123 135
124 235
134 145
234 245
125 345
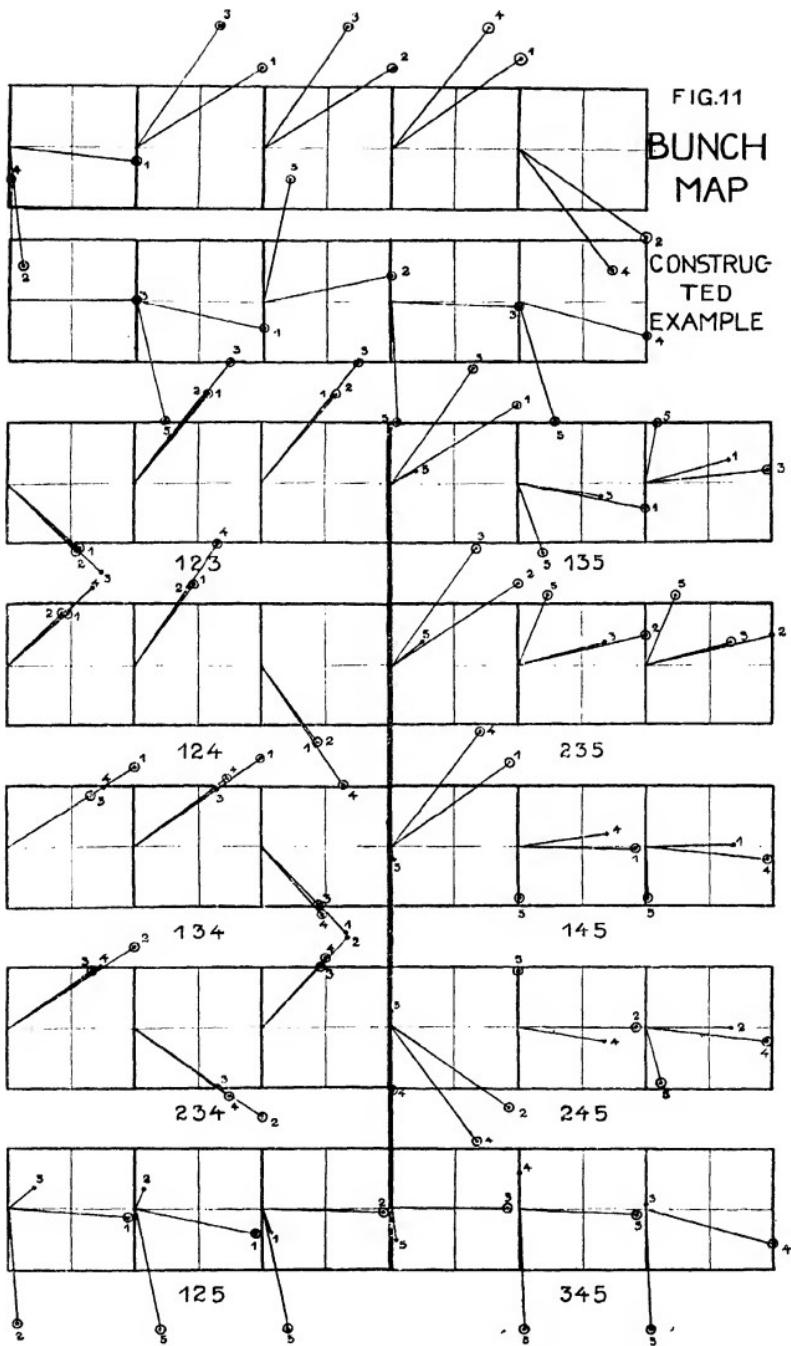
FIG. 11(CONTD.)

BUNCH MAP

CONSTRUCTED EXAMPLE

1234

1235

1245

1345

2345
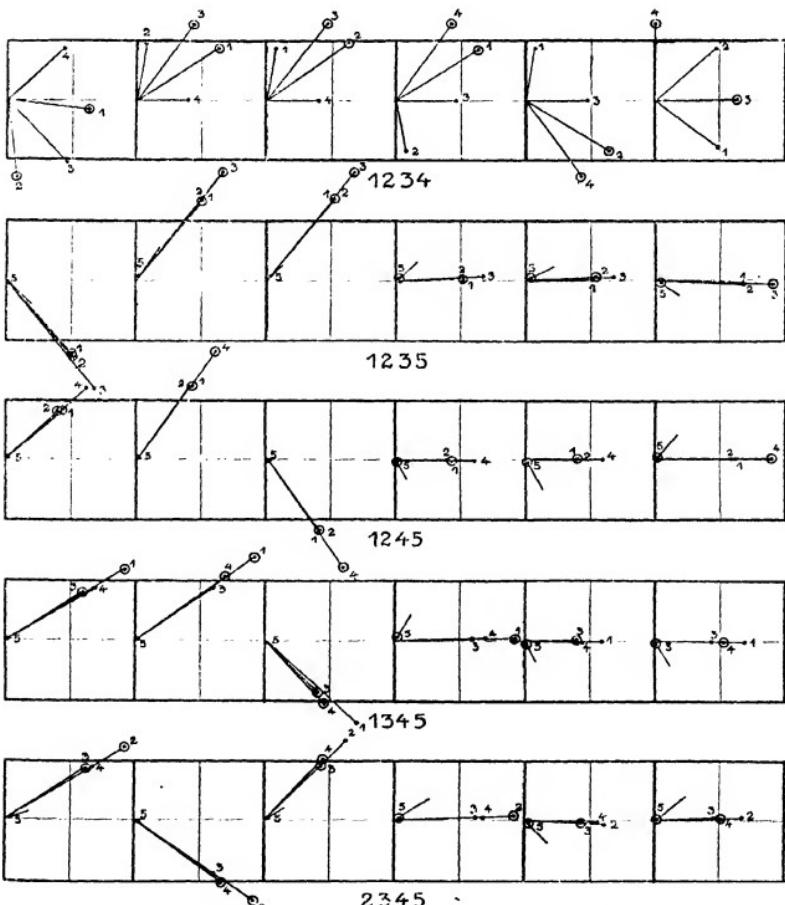
Errata: The (12) slope in (1235) has by an error in drawing been made about 23 pct. too steep.

Adding 5 to the set (12), we find that this variate is quite irrelevant. It does not tighten the bunch (it rather opens it), and the 5-beam lies outside and is much shorter than the other beams, and these other beams are not appreciably shortened by the inclusion of 5, nor are their directions changed.

Proceeding now to the four-sets, we see that in (1234) there is a definite *explosion* of the (12) bunch. It is particularly illuminating to watch the behaviour of the (12) bunch as we pass either from the set (123) or (124) to the bigger set (1234). In the former sets the tightness is excellent, while in the latter set there is hardly any organisation at all left in the bunch. 3 is decidedly *detrimental* for the (12) slope, when added to the set (124) and 4 is the same when added to (123).

It is also interesting to note the *degradation effect* produced in the (12) coefficient in the (1234) set. The beams 1 and 2 in this bunch *are virtually the same as in the original set* (12). Of course the absolute scale is different in the (1234) set, but the slopes of the beams are practically the same. The scale in the set (1234) is 40 times as large as in the other sets considered so far. This is a further indication that (1234)) is a multiply collinear set. Thus the consideration of the bigger set which includes both (123) and (124) verifies the suspicion we already got by studying these two sets separately.

If we add 5 either to (123) or to (124) the bunch of the (12) coefficient is virtually unchanged; a glance at the figure is sufficient to show that none of the beams are appreciably changed. Furthermore, in both cases the 5-beam is very much smaller: it is indeed so small that, compared with the other beams, it only appears as a tiny point near origin, all of which indicate 5 as superfluous in the determination of the (12) coefficient.

A similar conclusion is reached by adding 5 to the set (1234) and hardly any change is produced in the (12) bunch. The bunch in the set (12345) is just as exploded as it was in the set (1234).

Now consider the (13) coefficient. In the set (13) itself there is no significant tightness of the (13) bunch, but adding 2 or 4, we get immediately very tight bunches. In particular the (13) bunch in the set (134) is so close that, in the scale used in the figure, the three beams falls literally in one line. Again we see that the two (13) slopes obtained in the sets (123) and (134) are

quite different, which means that either they must belong to two different equations, or be very heavily biassed (i. e. not *net* determinations). A glance at the (13) bunch in the set (1234) verifies that we must have the former case: the bunch in the (1234) set is indeed distinctly exploded as compared with either the (123) or the (134) set. Adding 5 we find again that it is superfluous.

In a similar way we can discuss each of the other inter-coefficients (23), (14), etc. The result of the discussion is summarised in the star map in Figure 12. Here an asterisk, a circle or a blackball means a useful, superfluous and detrimental variate respectively.

On the basis of the star map and the bunch map we can now discuss the equations as such from the view-point of linear confluency, using the criteria of Section 18. A glance at the star map tells us immediately that all three-sets contained in (1234) are promising. Indeed, whatever variate we add, it appears as useful. And from the bunch map we see that each of these variates is not only useful but does what is necessary to produce a good fit. Indeed, *all* the bunches in the set (123) are good, and so are all the bunches in the set (124) etc. They are so good that it seems difficult to escape the conclusion that any of these sets is a closed, admissible set. The final test on this is given by the horizontal section (1234) in the star map (consisting of the four lines 1, 2, 3, 4). Each and all of the signs in this section is a blackball, indicating that no matter what variate we add in order to get the set (1234), it will appear as detrimental. And this applies no matter which one of the intercoefficients we consider. And taking a glance at the bunch map we see that this detrimental effects is violent: *all* the bunches in the set (1234) are definitely exploded.

It is also illuminating to study — by means of the star map — the four sets obtained by adding 5 to any of the three-sets contained in (1234). Take, for instance, the set (1235). All the variates 1, 2, 3 in this set are indicated exclusively by asterisks, while 5 is indicated exclusively by circles. Similarly in the sets (1245), (1345) and (2345). The superfluity of 5 is also checked by the zero slope criterion. Indeed, from the bunch map we see that the intercoefficient between 5 and any one of the variates 1, 2, 3, 4 is zero in all the four-sets containing 5.

The conclusion is thus definitive. Each of the three-sets

VARIATE ADDED	INTERCOEFFICIENT									
	12	13	23	14	24	34	15	25	35	45
1			*							
2	*	*								
3	*									
1					*					
2				*						
4	*									
1						*				
3				*						
4		*								
2						*				
3					*					
4			*							
1								*?		
2							○			
5	○									
1									*?	
3							○			
5		○								
2									*?	
3								*?		
5			○							
1										●?
4							●			
5				○						
2										○
4								●		
5					○					
3										○
4									○?	
5						○				
1			●		●	●				
2		●		●		●				
3	●	●		●	●					
4	●	●	●							
1			*					*?	*	
2		*					*	*		
3	*						*			
5	○	○	○							
1					*			*		*
2	*			*			*	*		*
4	*						*			
5	○			○	○					
1							*		*	*
3				*			*		*	*
4		*					*		*	
5		○		○						
2							*		*	*
3					*	*		*	*	*
4			*					*	*	
5			○		○	○				
1			●		●	●		○	○	○?
2		●		●		●	○	○	○	○?
3	●	●	●	●	●		○	○	○	○?
4	●	○	●						○	
5	○	○	○	○	◎	○				

contained in (1234) is a closed set, and hence (1234) itself multiply collinear, while 5 is a variate entirely extraneous to the system. This is just the conclusion we ought to find.

The set (12345) is primarily interesting because of the degradation effects and the persistency effects that manifest themselves here. The leading beams in the (12) coefficient in this all-inclusive set are just the same as they were in the original two-set (12), and the same applies to the inter-coefficients (13), (23), (14), (24), (34). This is the degradation effect. None of these bunches are tight; they are all exploded.

But the bunches (15), (25), (35), (45) are tight in the set (12345) even though this set is multiply collinear. More precisely these bunches show a zero-slope. The explanation is that, even in the all-inclusive set, these intercoefficients will have a meaning. No matter *how* we derive a five-variate equation from any of the closed admissible equations (123), (124) . . . (1235) . . . etc. the new equation will have the *same* coefficient of the variate No. 5, namely zero. This is the persistency effect discussed in Section 11. Incidentally, since the persistency effect here gives zero slope (instead of some other well defined slope), we can take it as an additional criterion of the superfluity of variate No. 5.

25. COMPATIBILITY SMOOTHING OF THE REGRESSION COEFFICIENTS IN THE CONSTRUCTED EXAMPLE.

Since all the three-sets contained in (1234) are admissible, we are confronted with a problem of compatibility smoothing of the regression coefficients in these 4 sets. For instance, if the coefficients in (123) are determined empirically as the diagonal regression coefficients in this set, and similarly in (124), we can from these results *derive* the coefficients that ought to exist in the set (134); and these may not coincide exactly with the diagonal regression coefficients determined directly in the set (134).

We shall first see how the compromise can be made by the method of diagonal zeros explained in Section 21. The matrix $A_{\kappa i}$ in (21.1) is in the present case determined by forming the $N = 4$ diagonal regression equations in the 4 three-sets contained in (1234). The material for this — when the variates are

taken in the normalised form — is of course contained in the tilling tables. Performing the rootsquaring we get

TABLE (25. 1)   REGRESSION COEFFICIENTS IN THE CONSTRUCTED EXAMPLE.

A_{Ki}	$i = 1$	2	3	4	Row sums (Upper number is the natural row sum, lower the absolute row sum)
(234) $K = 1$	.000000	.999896	— .680377	.753281	1 .072800 2 .433554
(134) 2	.999896	.000000	— .658590	— .754057	— .412751 2 .412543
(124) 3	.680377	— .658590	.000000	— .992585	— .970798 2 .331552
(123) 4	.753281	.754057	— .992585	.000000	.514753 2 .499923

Reducing by the absolute row sums we get

TABLE (25. 2) REGRESSION COEFFICIENTS REDUCED TO ABSOLUTE ROW SUM UNITY.

a_{Ki}	$i = 1$	2	3	4	Row sums (natural and absolute)
$K = 1$	.000000	.410879	— .279582	.309539	.440836 1.000000
2	.414457	.000000	— .272986	— .312557	— .171086 1.000000
3	.291813	— .282469	.000000	— .425719	— .416375 1.000001
4	.301322	.301632	— .397046	.000000	.205908 1.000000

The compatibility table defined by (21. 7) and computed on the basis of the figures in (25. 2) turns out to be

TABLE (25.3)   COMPATIBILITY TABLE.

$a_{Ki}^{(PQ)}$		$i=1$	2	3	4	Absolute row sum
	23	.000000	.415195	— .282520	.302285	1.000000
$K=1$	24	.000000	.414644	— .272980	.312377	1.000001
	34	.000000	.414915	— .277665	.307420	1.000000
	13	.418708	.000000	— .275785	— .305507	1.000000
2	14	.418294	.000000	— .266257	— .315449	1.000000
	34	.418504	.000000	— .271099	— .310396	.999999
	12	.289743	— .280463	.000000	— .429794	1.000000
3	14	.294609	— .275595	.000000	— .429796	1.000000
	24	.285035	— .285171	.000000	— .429794	1.000000
	12	.299326	.299634	— .401039	.000000	.999999
4	13	.304304	.294723	— .400973	.000000	1.000000
	23	.294177	.304718	— .401105	.000000	1.000000

In the actual work it will be found convenient to compute also an intermediate table giving the results of (21.6) (possibly with a sign factor). In this intermediate table both the natural and the absolute row sums are carried, while in (25.3) only the absolute row sum is carried.

A horizontal Section of (25.3) corresponds to a row in (25.2), and the coefficients in these two tables are directly comparable. For instance for $i=1$, $K=4$, the figures 0.299, 0.304, 0.294 are to be compared with 0.301. If the four regression equations had been exactly compatible, all these numbers would have been equal. The amount of spread which is actually present in these figures, indicate the degree of non-compatibility which exists between the equations used.

According to (21.8) we next form the simple arithmetic average of the three figures in each *(Ki)* cell of (25.3). Then according to (21.9) we take the average between the above average and the original figure in (25.2). This gives the figures on the rows $\nu=1$ in the various cells of (25.4)

TABLE (25.4) SUCCESSIVE SMOOTHING.

$a_{Ki}^{(\nu)}$		$i = 1$	2	3	4	Row sums (natural and absolute)
$K=1$	$\nu = 0$	0.000000	0.410879	— 0.279582	0.309539	.440836
						1.000000
	1	.000000	.412898	— .278652	.308450	.442696
						1.000000
	2	.000000	.412897	— .278652	.308451	.442696
						1.000000
2	$\nu = 0$	.414457	.000000	— .272986	— .312557	— .171086
						1.000000
	1	.416480	.000000	— .272017	— .311504	— .167041
						1.000001
	2	.416478	.000000	— .272017	— .311505	— .167044
						1.000000
3	$\nu = 0$	.291813	— .282469	.000000	— .425719	— .416375
						1.000001
	1	.290804	— .281440	.000000	— .427757	— .418393
						1.000001
	2	.290805	— .281440	.000000	— .427755	— .418390
						1.000000
4	$\nu = 0$	.301322	.301632	— .397046	.000000	.205908
						1.000000
	1	.300295	.300662	— .399043	.000000	.201914
						1.000000
	2	.300296	.300662	— .399042	.000000	.201916
						1.000.000

These figures we now take as a new starting point. Going through exactly the same process once more we find as a second smoothing the figures on the row $\nu = 2$ in the various cells of (25.4). To give an idea of the extreme rapidity with which the process converges towards a situation where all the 4 equations are exactly compatible I give the compatibility tables for $K = 4$ computed with 10 decimal places, and the

similar tables for the first and second smoothing[1]. In these tables are added for comparison also the coefficients that served as the starting point for each smoothing. These are the same as those given in (25.4)

TABLE (25.5)   COMPATIBILITY TABLE FOR THE UNSMOOTHED COEFFICIENTS.

		$i=1$	2	3	4	Absolute row sum
$K=4$	$(\nu=0$ in (25.4))	.3013216807	.3016320903	— .3970462290	.0000000000	1.0000000000
	12	.2993261512	.2996345051	— .4010393437	.0000000000	1.0000000000
	13	.3043031139	.2947220753	— .4009748108	.0000000000	1.0000000000
	23	.2941775623	.3047163358	— .4011061020	.0000000000	1.0000000001

TABLE (25.6)   COMPATIBILITY TABLE FOR COEFFICIENTS SMOOTHED ONCE.

		$i=1$	2	3	4	Absolute row sum
$K=4$	$(\nu=1$ in (25.4))	.3002953116	.3006615312	— .3990431573	.0000000000	1.0000000001
	12	.3002965468	.3006626180	— .3990408351	.0000000000	.9999999999
	13	.3002923084	.3006667436	— .3990409480	.0000000000	1.0000000000
	23	.3003009316	.3006583501	— .3990407183	.0000000000	1.0000000000

TABLE (25.7)   COMPATIBILITY TABLE FOR COEFFICIENTS SMOOTHED TWICE.

		$i=1$	2	3	4	Absolute row sum
$K=4$	$(\nu=2$ in (25.4))	.3002959536	.3006620509	— .3990419956	.0000000000	1.0000000001
	12	.3002959536	.3006620509	— .3990419955	.0000000000	1.0000000000
	13	.3002959537	.3006620509	— .3990419955	.0000000000	1.0000000001
	23	.3002959535	.3006620509	— .3990419956	.0000000000	1.0000000000

For $K=4$, $i=1$, for instance, we see that the difference between the highest and lowest of the unsmoothed coefficients which ought to be compatible is about 0.01. Already by the first smoothing this difference reduces to about 0.000009 and in the second smoothing it virtually does not show up in the ten decimal places carried.

[1] The original regression coefficients were computed with 6 decimal places; for the purpose of the calculation here considered four zeros were added throughout. For the experimental verification of the rapidity of convergency this is just as good as if we had added any other figures.

As an illustration we shall also show how the compatibility smoothing by the general method of Section 22 is carried out.

The matrix (25. 1) reduced to row sum square unity [1] is given in (25. 8).

TABLE (25. 8) REGRESSION COEFFICIENTS REDUCED TO ROW SUM SQUARE UNITY.

a_{Ki}	$i = 1$	2	3	4	Row sums (natural and square row sums)
$K = 1$	.000000	.701766	— .477515	· 528682	0.752933 1.000001
2	.706657	.000000	— .465446	— .532915	— 0.291704 1.000002
3	.495968	— .480086	.000000	— .723556	— 0.707674 1.000000
4	.517202	.517735	— .681508	.000000	0.353429 1.000001
Natural Column sums	1.719827	.739415	— 1.624469	— .727789	0.106984

The row moments of (25. 8) are

TABLE (25. 9) ROW MOMENTS OF (25. 8).

μ_{KH}	$H = 1$	2	3	4	Natural row sums
$K = 1$	1.000000	— 0.059485	— 0.719439	0.688759	0.909835
2		1.000000	0.736072	0.682689	2.359276
3			1.000000	0.007958	1.024591
4				1.000000	2.379406

In the present case the number of independent regressions is $\varkappa = 2$. We consequently need to till the matrix (25. 9) up to and including the three rowed sets. The result is given in (25. 10) and (25. 11)

---

[1] The symbol a_{Ki} in (25. 8) is not exactly the same as in (25. 2) because the row reductions are now by square sums instead of absolute sums.

TABLE (25.10)   TWO-ROWED TILLING TABLES.

	1	2		2	3		2	4
1	1.000000	.059485	2	1.000000	− .736072	2	1.000000	− .682689
2		1.000000	3		1.000000	4		1.000000
△ =	.996462	.996462	△ =	.458198	.458198	△ =	.533936	.533936

	1	3		1	4		3	4
1	1.000000	.719439	1	1.000000	− .688759	3	1.000000	− .007958
3		1.000000	4		1.000000	4		1.000000
△ =	.482408	.482408	△ =	.525611	.525611	△ =	.999937	.999937

TABLE (25.11)   THREE-ROWED TILLING TABLES.

	1	2	3
1	.458198	− .470074	.675654
2		.482408	− .693276
3			.996462
△ =	.000069	.000069	.000069

	1	3	4
1	.999937	.724920	− .694484
3		.525611	− .503478
4			.482408
△ =	.000069	.000069	.000069

	1	2	4
1	.533936	.529693	− .729369
2		.525611	− .723659
4			.996462
△ =	.000068	.000069	.000069

	2	3	4
2	.999937	− .730639	− .676831
3		.533936	.494550
4			.458198
△ =	.000069	.000069	.000069

The cumulated matrix defined by (22.9) is given in (25.12)

TABLE (25.12)   CUMULATED MATRIX.

$\mu_{(K\,H)}$	$H = 1$	2	3	4
$K = 1$	1.992071	.059619	1.400574	− 1.423853
2		2.007956	− 1.423915	− 1.400491
3			2.056009	− .008928
4				1.937068

Sum of diagonal elements = 7.993104   } (checked by (22.10))
Sum of elements in north east triangel = − 2.796994

On the basis of the figures in (25.8) and (25.9) the averages $\bar{a}_{Ki}$ defined by (22.7) becomes as in (25.13). This table is checked by row sums and column sums. (The column and row sums are also needed for subsequent checks).

TABLE (25. 13).

$\overline{a}_{Ki}$	$i=1$	2	3	4	Natural row sum
$K=1$	— .000176	.707593	— .473185	.524663	.758895
2	.712442	— .000178	— .461154	— .528797	— .277687
3	.491650	— .475802	— .000022	— .729221	— .713395
4	.513196	.513624	— .687515	— .000019	.339286
Natural Column sum	1.717112	.745237	— 1.621876	— .733374	.107099

The matrix c_{Ki} defined by (22. 13) is

TABLE (25. 14).

c_{Ki}	$i=1$	2	3	4	Natural row sum
$K=1$	.000176	— .005827	— .004330	.004019	— .005962
2	— .005785	.000178	— .004292	— .004118	— .014017
3	.004318	— .004284	.000022	.005665	.005721
4	.004006	.004111	.006007	.000019	.014143
Natural Column sum	.002715	— .005822	— .002593	.005585	— .000115

And the matrices ν_{KH} and γ_{KH} defined by (22. 15) and (22. 16) are

TABLE (25. 15).

ν_{KH}	$H=1$	2	3	4	Natural row sum
$K=1$	.000206	— .000005	— .000045	.000052	.000208
2		.000208	.000047	.000050	.000300
3			.000198	— .000001	.000199
4				.000214	.000315

Grand total .001022

TABLE (25. 16).

γ_{KH}	$H=1$	2	3	4	Natural row sum
$K=1$	.000069	.000000	.000048	— .000049	.000068
2		.000069	— .000049	— .000048	— .000028
3			.000069	.000000	.000068
4				.000069	— .000028

Grand total .000080

The grand totals of (25.15) and (25.16) are checked by (22.17) and (22.18) respectively.

The coefficients of the subcharacteristic polynomials $F_K(\lambda)$ ($K = 1, 2, 3, 4$), (that is the first principal minors in (22.19)), can in this case be determined directly by (22.22), each determinant being computed directly by Sarrus rule. All the work can be done in one stroke if one has a multiplication machine with an extra arrangement for grand total (besides sub-total) cumulation. The same can also be done — although not quite as easily — if the operator uses at the same time an ordinary multiplication machine and a listing adding machine. It is convenient to write (or better typewrite) the columns of the matrices μ, ν and γ on loose strips and permute them so as to obtain the various terms in (22.22). It will be found that all the terms vanish (at least in the first 6 decimal places) except those that contain two affixes μ. These terms are given in (25.17).

TABLE (25.17) COEFFICIENTS OF SUBCHARACTERISTIC
POLYNOMIALS $F_K(\lambda)$

	F_1	F_2	F_3	F_4
$S'_{\mu\mu\mu}$	.000069	.000069	.000068	.000069
$S_{\mu\mu\nu}$	.000275	.000277	.000278	.000271
$S_{\mu\mu\gamma}$	.000274	.000276	.000283	.000268

In view of the fact that the other terms S vanish, each column in (25.17) gives directly the coefficints of the polynomial F_K in question (with the sign of the middle term changed, see (22.21)).

The computation of these coefficients is much simpler than may appear from the description. As a matter of fact it is so simple that there is no use applying any check during the work. But the final result must of course be checked and this is done most conveniently by computing the values of the four polynomials F_K for some value of λ, say $\lambda = 0.1$, according to the formulae obtained, and verifying that these values are equal to those obtained by evaluating the corresponding three-rowed determinants directly after having inserted this value of λ. In the present case this actually checked immediately for all the four polynomials. The following values were found:

TABLE (25. 18)   VALUES OF THE SUBCHARACTERISTIC POLYNOMIALS
for $\lambda = 0.1$.

	Computed by (25. 17)	Computed by evaluation of the determinant
F_1	0.000044	0.000043
F_2	0.000044	0.000045
F_3	0.000043	0.000042
F_4	0.000045	0.000044

Since by the argument in connection with Figure 10 of Section 22 all the subcharacteristic polynomials must touch zero, (not pass through zero), the two roots of any of the second order polynomials obtained ought to coincide. This is actually verified, the zeros being as indicated in (25. 19).

TABLE (25. 19)   ZEROS OF SUBCHARACTERISTIC POLYNOMIALS;
(EACH ZERO IS DOUBLE).

For F_K	λ
$K = 1$	0.498
2	0.502
3	0.491
4	0.506

The fact that all the zeros given in the second column of (25. 19) nearly coincide shows that it is possible to select λ in (22. 12) in such a way that the smoothed matrix a'_{Ki} has a row moment matrix (22. 14) which comes very near to having simultaneously all its principal minors of order $\varkappa + 1 = 3$, equal to zero. This smoothed a'_{Ki} will then be very near to fulfilling the compatibility condition we want to satisfy.

The average value of the four zeros in (25. 19) is for practical purposes *exactly* 0.5. In other words the present more elaborate analysis leads to adopting *just the simple kind of average which was suggested heuristically in (21. 9)*. Adopting this average, we get the following smoothed values.

TABLE (25. 20).

a'_{Ki}	$i = 1$	2	3	4	Natural row sum
$K = 1$	— .000088	.704679	— .475350	.526672	.755913
2	.709549	— .000089	— .463300	— .530856	— .284696
3	.493809	— .477944	— .000011	— .726388	— .710534
4	.515199	.515679	— .684511	—. 000009	.346358

If wanted, the smoothing can be repeated, but in the present case this will hardly be worth while.

In order to get equations that can be directly compared with those found previously we ought to combine the coefficients (25. 20) in such a way as to reduce the diagonal elements to zero. Since the coefficients of (25. 20) are not yet *exactly* compatible, it may have a small effect *how* the elimination is performed. Doing it alternately by the various equations and taking the average we get

TABLE (25. 21).

	$i = 1$	2	3	4
$K = 1$	.000000	.704680	— .475410	.526642
2	.709538	.000000	— .463359	— .530790
3	.493800	— .477952	.000000	— .726388
4	.515194	·515685	— .684510	.000000

The results thus obtained we may now compare with the coefficients obtained by using the "true" regressions (13. 3) by which the data were constructed. Since the empirical regression coefficients determined by the preceding methods are worked out in normalised coordinates, we must either transform (13. 3) to normalised coordinates, or transform the final values in (25. 4) (for $\nu = 2$) and (25. 21) to non-normalised coordinates. We prefer the former. This means that the coefficients of (13. 3) must be reduced by the square roots of the diagonal elements in (13.8). Doing this we find that the four "true" regressions in normalised coordinates are

TABLE (25. 22).

	$i = 1$	2	3	4	Natural row sum
$K = 1$	.000000	2.013854	— 1.331401	1.514070	2.196523
2	1.993566	.000000	— 1.331401	— 1.514070	— .851905
3	.996783	— 1.006927	.000000	— 1.514070	— 1.524214
4	.996783	1.006927	— 1.331401	.000000	.672309

Reducing this to absolute row sum and row sum square equal to unity respectively, we obtain:

TABLE (25. 23.) »TRUE» REGRESSIONS REDUCED TO ABSOLUTE ROW SUM 1.

	$i = 1$	2	3	4	Row sums (natural and absolute)
$K = 1$	.000000	.414431	— .273989	.311580	.452022 1.000000
2	.411976	.000000	— .275138	— .312887	— .176049 1.000001
3	.283356	— .286239	.000000	— .430405	— .433288 1.000000
4	.298876	.301917	— .399207	.000000	.201586 1.000000

TABLE (25. 24) »TRUE» REGRESSIONS REDUCED TO ROW SUM-SQUARE 1.

	$i = 1$	2	8	4	Row sums (natural and square)
$K = 1$	.000000	.706696	— .467211	.531313	.770798 1.000000
2	.703107	.000000	— .469569	— .533994	— .300456 1.000000
3	.480698	— .485590	.000000	— .730159	— .735051 1.000000
4	.512683	.517900	— .684789	.000000	.345794 1.000000

Comparing finally the last values (i. e. for $\nu = 2$) in (25. 4) with (25. 23) and (25. 21) with (25. 24), we obtain the following percentage errors:

TABLE (25. 25) PERCENTAGE DEVIATION OF EMPIRICAL REGRESSION COEFFICIENTS FROM »TRUE» VALUES.

(Compatibility smoothing by the method of Section 21).

	$i = 1$	2	3	4
	Pct.	Pct.	Pct.	Pct.
$K = 1$	.00	— .37	1.70	1.00
2	1.09	.00	— 1.13	.44
3	2.63	— 1.68	.00	— .62
4	.48	— .42	— .04	.00

TABLE (25. 26) PERCENTAGE DEVIATIONS OF EMPIRICAL REGRES-
SION COEFFICIENTS FROM ›TRUE› VALUES.

(Compatibility smoothing by the method of Section 22).

	$i = 1$	2	3	4
	Pct.	Pct.	Pct.	Pct.
$K = 1$	.00	— .29	1.75	.89
2	.91	.00	— 1.34	.60
3	2.73	— 1.60	.00	— .52
4	.49	— .43	— .04	.00

It is seen that the empirically determined coefficients come very close to the "true" values. On the average the error only amounts to a fraction of one per cent. Only in a few cases does the error run as high as one or two per cent.

It is further seen that the two methods lead essentially to the same result. In practice it therefore seems advisable to use whenever possible the simple method of 21. In most cases one single smoothing will probably be sufficient. Furthermore in practice one will as a rule not need to carry as many decimal places as we have used — for illustration purposes — in this and the preceding Section. The actual work will therefore be comparatively easy.

26. AN EXAMPLE IN 6 VARIATES FROM AMERICAN CONSUMPTION STATISTICS. MEASUREMENTS OF THE MONEY FLEXIBILITY.

In static economic theory one studies how the individual (the family) distributes its resources under a given system of prices and a given income. By making certain assumptions about the manner in which the cost of living enters into this mechanism one concludes that for a given commodity which is not in substitution connection with other commodities, there ought to exist the following equation [1]:

$$(26. 1) \qquad\qquad w(r) = \alpha \cdot u(x)$$

where $r =$ real (deflated) income.

         $x =$ quantity consumed per unit of time of the reference commodity.

[1] See for instance the present author's book ›New Methods of Measuring Marginal Utility›. Tübingen 1931.

(26. 2) $\qquad \alpha = \dfrac{P}{p} =$ inverted relative price of the reference com modity, p being its absolute price and P the cost of living.

The function $w(r)$ in (26. 1) is the function that exhibits how w, the marginal utility of the money unit, varies as a function of the income, and the function $u(x)$ represents how u, the marginal utility of the commodity unit, varies as a function of the quantity consumed. The equation (26. 1) defines a surface in(α, r, x) coordinates, the so called "surface of consumption". If a set values of (α, r, x) are observed they ought to lie on this surface. A statistical (α, r, x) scatter should therefore give us some information of the actual shape of the surface.

The money flexibility, namely

(26. 3) $\qquad \check{w} = \dfrac{d \log w(r)}{d \log r} = \dfrac{d w(r)}{d r} \cdot \dfrac{r}{w(r)}$

can be determined if the shape of the consumption surface is known. It can even be determined if only *a section* of the surface is known, namely one for which $x =$ constant. Such an (α, r) curve is called an *isoquant*. From (26. 1) we see that the logarithmic derivative of r with respect to α along any isoquant gives the flexibility. This forms the basis of the flexibility measurements I undertook in 1923 and 1930, and it also served as the basis for further work done by Frederick V. Waugh, Maurice H. Belz and myself during these gentlemen's stay in Oslo.

This further work was done on American, Swedish and Norwegian data of different kinds, household budgets as well as time series of national income and consumption. A complete account of the results obtained by this comparative analysis will be published separately as another of the Oslo Institute's publications. In the present connection I only select two sets of data that may examplify the use of the confluence technique discussed in the present paper.

The data in question were collected and prepared for this analysis by Dr. Waugh. It contains information which permits to compute the three variates (26. 2) for each year for the United States as a whole. Dr. Waugh wanted to apply the method to data relating to a whole country. In his mind the

flexibility measurement thus obtained would be more interesting than flexibility measurements referring to particular groups (customers in cooperative chain stores as in my study of 1923, or working men's families as in my 1930 study). Dr. Waugh also suggested that the rate of change with respect to time of α and r may exert some influence on the flexibility, so that the year to year changes of these two variates ought to be included in the analysis as new variates. This gave the analysis a more dynamic character. Finally, time itself was taken in as a catch all for the trend factors. The list of variates being thus enlarged some simplifications had to be made in the theoretical scheme. In my previous studies I had used methods that did not assume any particular form of the functions $w(r)$ and $u(x)$. We now decided tentatively to work with functions that were linear in the logarithms of the variates. This led to considering a linear regression between some or all of the following six variates.

$$x_1 = \log \alpha$$
$$x_2 = \log r$$
$$x_3 = \dot{x}_1 = \text{year to year change of } x_1$$
(26.4) $$\quad x_4 = \dot{x}_2 = \text{year to year change of } x_2$$
$$x_5 = \text{time (linear trend factor).}$$
$$x_6 = \log x \text{ (x being quantity consumed per year. Provisoric means were used for the numerical work).}$$

The intercoefficient of the variate No. 1 and the variate No. 2 in the regression equation connecting these variates will give the money flexibility (when the variates are taken in non-normalised coordinates).

The period which will be treated here is the post war period 1919—31, and the data will be used for the two reference commodities meat and butter for the United States. (In the complete work also other periods and commodities were considered). The correlation matrices were

TABLE (26.5). CORRELATION COEFFICIENTS IN MEAT
(1919—31. U. S.).

r_{ij}	$j = 1$	2	3	4	5	6
$i = 1$	1.000000	− 0.549342	0.212052	0.421796	− 0.664092	0.511130
2		1.000000	− 0.753343	0.327788	0.680848	0.402229
3			1.000000	− 0.534176	− 0.378352	− 0.497311
4				1.000000	− 0.203932	0.785384
5					1.000000	− 0.134523
6						1.000000

TABLE (26. 6). CORRELATION COEFFICIENTS IN BUTTER
(1919—31. U. S.).

r_{ij}	$j = 1$	2	3	4	5	6
$i = 1$	1.000000	− 0.070778	0.558808	− 0.217851	0.548585	0.513617
2		1.000000	− 0.460509	0.327788	0.680848	0.775721
3			1.000000	− 0.701721	0.122482	− 0.130910
4				1.000000	− 0.203920	0.265165
5					1.000000	0.788472
6						1.000000

The sumsquares of the variates 1 and 2 were

$$
\begin{array}{ll}
& \textit{Meat} \qquad\qquad\qquad \textit{Butter} \\
(26.7) & m_{11} = 0.016865 \qquad m_{11} = 0.025360 \\
& m_{22} = 0.020824 \qquad m_{22} = 0.020824
\end{array}
$$

These values are needed in order to get back from the nor-malised to the non-normalised variates.

From the data in (26.5) and (26.6) a complete tilling was done, and the bunch map for the intercoefficient (12) was drawn. Only the bunch map for this intercoefficient was considered, since the main object of the analysis was to investigate the money flexibility. The (12) bunch maps for meat and butter are given in Figures 13 and 14. It will be noted that these maps also contain a variate No. 7. This variate will be discussed later. For the moment we shall consider only that part of the maps that refer to the variates Nos. 1—6.

The bunches in Fig. 13 and 14 are drawn on scales as indigated in the following tables:

SCALES USED IN FIG. 13, MEAT.

Set	Scale of enlargement	Set	Scale of enlargement	Set	Scale of enlargement
12	1	1246	2	123467	10
123	»	12347	»	123456	20
124	»			123457	»
125	»	12345	5	123567	»
126	»	12346	»	124567	»
127	»	12356	»		
		12456	»	1234567	100
1234	»	12357	»		
1235	»	12457	»		
1245	»	12367	»		
1236	»	12467	»		
1256	»	12567	»		
1237	»				
1247	»				
1257	»				
1267	»				

SCALES USED IN FIG. 14, BUTTER.

Set	Scale of enlargement	Set	Scale of enlargement	Set	Scale of enlargement
12	1	1256	2	12567	20
123	»	12347	»	123456	»
124	»			123457	50
125	»	1257	5	123467	»
126	»	1267	»		
127	»			123567	100
		12345	»	124567	»
1234	»	12346	»		
1235	»	12356	»	1234567	500
1245	»				
1236	»	12456	10		
1246	»	12357	»		
1237	»	12457	»		
1247	»	12367	»		
		12467	»		

We shall discuss the meat map first. Looking at the three-sets we are immediately struck by the fact that it is the set (126) that has the best tightness. This is already a first indication of the plausibility of the theoretical basis which we started from, namely that there exists a structural relation

FIG. 13.

BUNCH MAP - MEAT
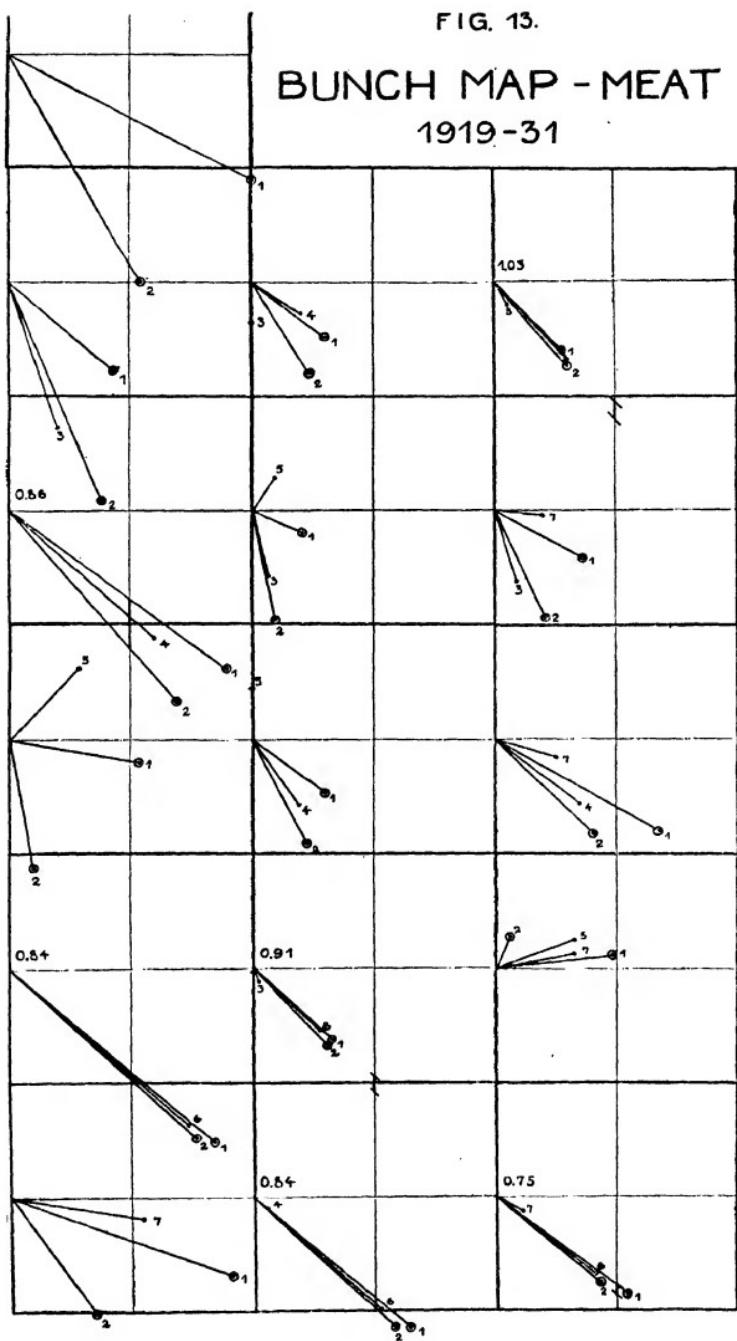
1919-31

FIG. 13 (CONTD.)

BUNCH MAP- MEAT

1919 – 31

FIG. 14
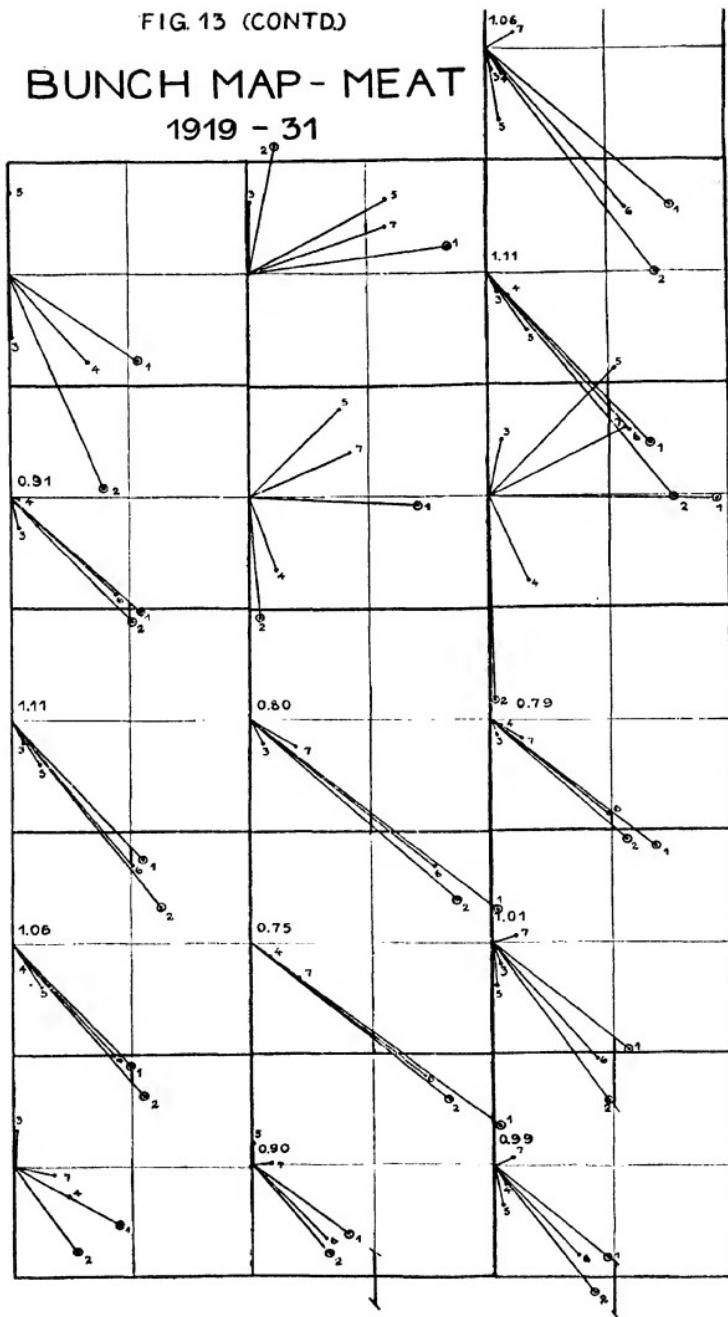
BUNCH MAP - BUTTER
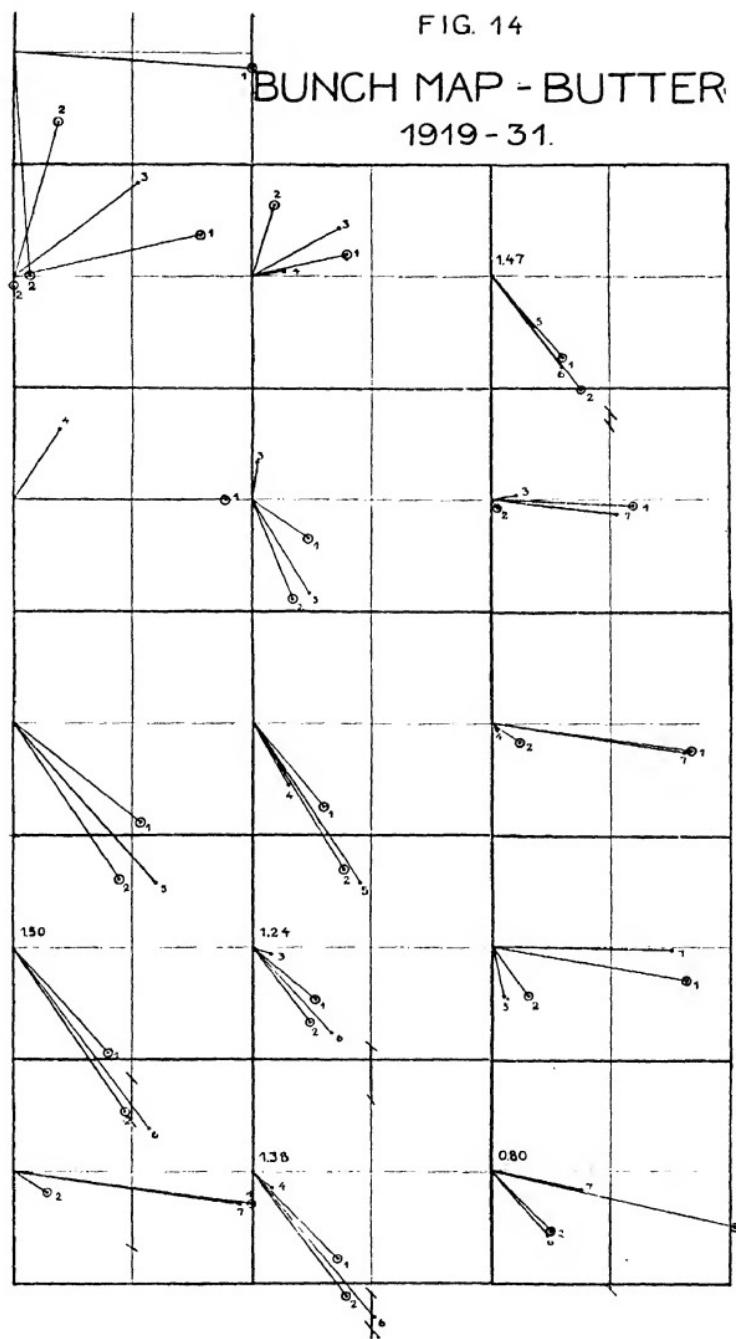
1919 - 31.

FIG.14 (CONTD.)

BUNCH MAP - BUTTER
1919 - 31.
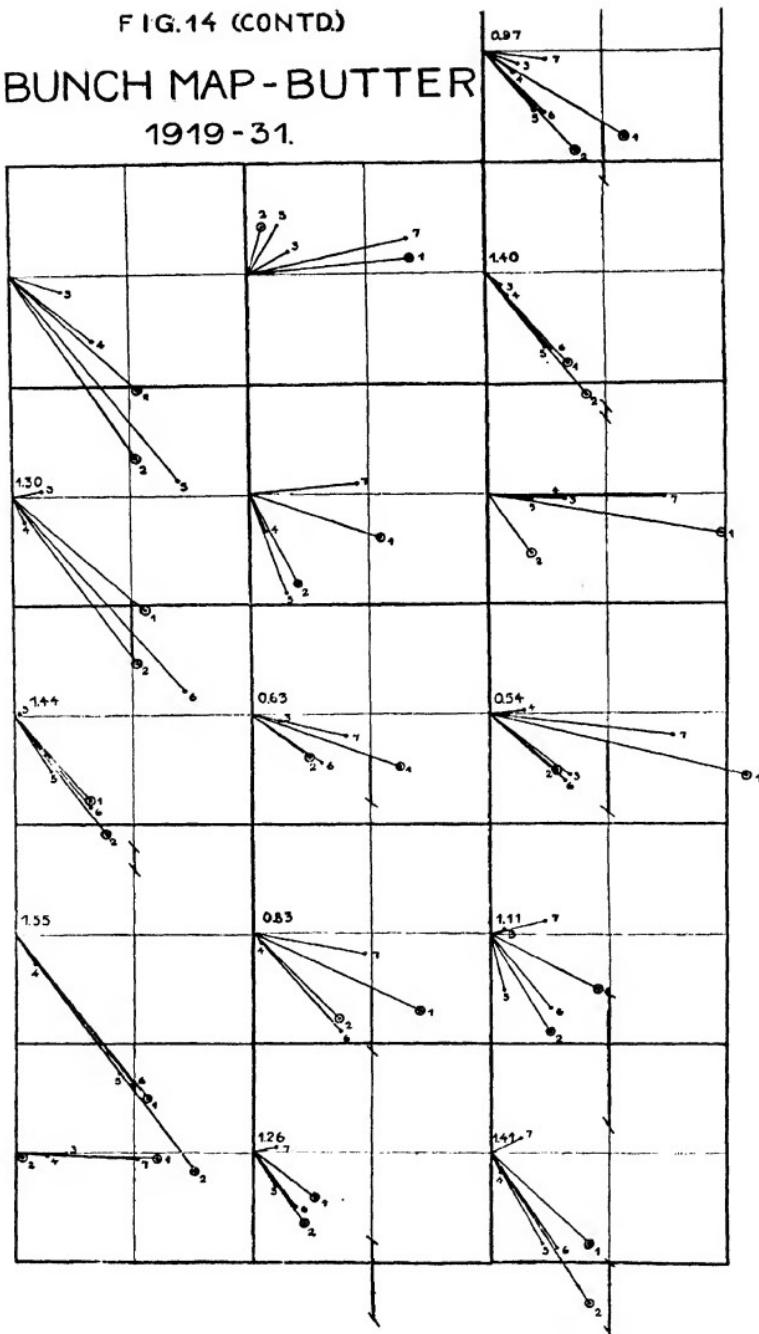
between the variates (126). Comparing the bunch in (126) with
that in (12), we see in particular that no statistically significant
information about the intercoefficient in question is obtained
before we add 6. This is a general feature which will be noted
also in all the other sets. Notice for instance how all the bun-
ches (12), (123), (124), (125), (1234), (1235), (1245), (12345) are
tightened when 6 is added. This gives an interesting empirical
corroboration of the "isoquant" idea (it will be remembered that
6 denotes the quantity consumed). The money flexibility as
computed by the diagonal regression in the set (126) is [1] 0.84.
This flexibility is found simply by de-normalising the diagonal
regression coefficients by means of (26. 7). The resulting
flexibilities are indicated in the map.

Now let us proceed to the four-rowed sets to test Dr. Waugh's
suggestion that some of the other variates may exert an in-
fluence. A glance at the map is sufficient to throw out all the
four-sets that do not contain 6. This leaves us with the 3 four-
sets (1236), (1246), (1256), in other words, the 3 sets obtained
by adding to (126), 3, 4 and 5 respectively. In all these sets
there is a relatively good tightness. The fit is perhaps slightly
inferior to that in (126), but there is no "explosion".

It thus seems to be allowed to add any of the variates 3, 4
or 5 to (126). The effect on the general slopes of the bunches
produced by the inclusion of these variates is as follows:
3 makes the slope a little steeper (flexibility 0.92 as against
0.84). 4 leaves it unchanged and 5 makes it a good bit steeper
(flexibility 1.03 as against 0.84). The general slopes of the
bunches, as well as their tightness, are seen most easily from
the small vertical bars that indicate the distance between the
prolongation of the leading beams [2]. The steepening effect of 5
is quite significant since the (1256) sector lies entirely outside
(on the lower side) of the (126)-sector. This is interesting: it
means that the trend connection between the variates has —
by being left in the material — biassed the result observed in
the set (126).

It seems probable that the difference in slope, say in the sets
(1246) and (1256) is not due to these being slopes in two

---

[1] In the following I say for brevity that the flexibility is 0.84, 1.03, etc.
meaning by this that the flexibility is — 0.84, — 1.03, etc. All the flexibilities
considered are negative; indeed, all the bunches considered are sloping down.

[2] The leading beams are marked with small circles.

independent equations, but due only to the bias produced by neglecting to eliminate simultaneously the effects of the various variates. In other words it appear to be a gross slope effect not a multilinear effect. In order to test this we want to see if there is produced an explosion when we go to the higher sets.

Looking at the five-sets we see again that there can be no question of accepting a set where 6 is not included. This leaves us with the 3 sets (12346), (12356) and (12456), that is with the sets obtained by adding (34), (35) and (45) respectively to (126). All these sets are still fairly good; (12356) gives the steepest slope (flexibility 1.11) which was to be expected since in the four-sets both 3 and 5 had a tendency to steepen the slope. In (12356) these steepening effects will be accumulated. Again we find that 4 virtually does not change the slope. The flexibility is 0.91 in (12346) as against 0.92 in (1236) and 1.06 in (12456) as against 1.03 in (1256).

The flexibility in the 3 five-sets considered are somewhat different, namely 0.91, 1.11 and 1.06 respectively, and the tightness of the slopes are such that these differences appear significant. We are therefore again in the situation that it is desirable to test the higher sets. Only one such set exists, namely the one including all the six variates considered. Even in this set the tightness is fairly good, indicating that there is no danger in considering this set, the tightness is even better than in most of the subsets. In this big set the slope is the same as in the five-set that included both the steepening factors 3 and 5. In both cases the diagonal flexibility is 1.11. This is another verification of the fact that 4 does not influence the result.

It is very instructive to look at the appearance of the whole bunch in (123456). The various beams in this total bunch give an excellent expression for the importance of the various variates and for the sense in which they will bias the (12) slope if their influence is not taken account of by including them in the regression equation. In the first place we notice that (126) are the three important variates. Their beams are much longer than the other beams in the big set (123456). And the relative disposition of the three important beams remains practically the same in all lower bunches down to the set (126). The only effect of the other variates is that the sub-bunch of (126) is swayed a little up or down according to which supplementary

variate (-s) we include. We have seen that 4 had practically
no effect all the way through. This is summarised by the posi-
tion of the 4-beam in the set (123456). Indeed, 4 is a very
short beam: it has a small power to "enforce its will". And
furthermore its direction coincides with the average slope of
the leading beams: it "wants to go the same way" as the beams
of the two variates whose intercoefficient we are discussing.
Both these facts make it plausible that 4 does not influence the
result observed. On the other hand the variates 3 and 5 were
observed to exert a steepening influence all the way through,
the effect from 5 being the strongest. This is also summarised
in the bunch of the big set (123456) by the position of the
beams 3 and 5. Indeed, both are steeper than the leading slope
(they try to "pull down"), and 5 is the longer of the two.

When all these facts are taken into consideration, we may
formulate the following conclusion. If no disturbing factor
outside the set (123456) is taken into account, the money
flexibility computed by using meat as a reference commodity is
about 1.1, the first decimal place after the comma in this
magnitude is probably significant. At least it seems to be
beyond doubt that when the effects of the variates indicated are
eliminated, the flexibility is larger than unity. Using the
significance factor (19. 4) we may express the result in the form

$$(26.8) \qquad -\breve{w} \; \frac{\text{(via meat)}}{\text{(123456)}} = 1.11 \cdot / \cdot \; 0.94$$

This is a figure considerably higher than the average of the
flexibilities I found in my 1930 study using United States
budget data. The difference has an interesting explanation
which will suggest itself when we have subjected butter to a
similar analysis to that which was just carried through for
meat.

The bunch map for butter is given in Figure 14. Here one
will immediately be struck by the fact that in the *lower* sets
the tightness is poorer than it was in meat, although also butter
is fairly good even in the lower sets provided 6 is included. The
need for including 6 is just as marked in butter as it was in
meat, and the reader may himself verify how the inclusion of
6 is everywhere the important thing that brings order into the
matter. This is a further corroboration of the "isoquant" idea.

Another striking feature of the map is that the money

flexibility measured with butter as a reference commodity seems to give a substantially higher value than we got when meat was used. Disregarding the changes produced by taking in one or the other of the supplementary variates, we may say roughly that the money flexibility measured via butter lies around 1.4 or 1.5, while measured via meat we found it to be 1.1.

Looking now at the details of the map we see that, adding *any* of the variates 3, 4 or 5 to (126), we get an improvement in tightness. Therefore, any of these variates must be considered useful.

But the effects of their inclusion are not the same, 3 is definitely flattening the slope (the flexibility in (1236) is 1.24 as against 1.5 in (126)), 4 is also flattening the slope, but not so much (flexibility in (1246) 1.38 as against 1.5 in (126)), while 5 has hardly any effect on the slope; 5 has, however, a very strong effect on the tightening. If we would only think of the tightening effect, we would conclude that it is more important to add 5 than 3, but, so far as the correct determination of the *net* flexibility slope is concerned, it is more important to add 3 than 5. This is another example of how useful it is to work *simultaneously* with slope and tightness criteria as we do in the bunch map analysis (as distinct from the analysis by the test parameters of Part II, which only express tightness).

The difference in results thus obtained in the four-sets by adding 3, 4 or 5, makes it necessary to consider the fivesets to see whether the difference in slope observed in the four-set was a gross-slope effect or a multilinear effect.

Amongst the five-sets, one, namely (12456), stands out. The tightness in this bunch is excellent. This is highly significant. Indeed, 4 and 5 were the two variates that produced most tightening when added to (126). The fact that the tightening is still better when *both* 4 and 5 are added, means that the difference in the slopes in (1246) and (1256) was a gross-slope effect, not a multilinear effect. A similar conclusion is reached in the set (12356), while in (12346) the situation is doubtful. We need not however worry much about (12346). Already a comparison between (12356) and (12456) shows that we are again confronted with the same question as to whether the difference in slopes in these two sets (flexibility 1.44 in the

former and 1.55 in the latter), shows gross slope effect or a multilinear effect. We therefore need to consider also the six set (123456). *This big set actually shows the most perfect tightness of all the sets considered.* This applies not only to the leading beams, but also to the secondary beams. There can therefore be no doubt that this is the set to be retained. It gives a flexibility of

$$(26.9) \qquad -\check{w} \; \frac{\text{(via butter)}}{\text{(123456)}} = 1.40 \; \cdot / \cdot \; 0.96$$

The whole appearance of the bunch (123456) verifies again the conclusions we reached *en route*: 3 is flattening and 4 too but not so much, while 5 is hardly influencing the slope.

27. SUBSTITUTION POSSIBILITIES IN MEAT AND BUTTER. THE INCLUSION OF THE ABSOLUTE PRICE BESIDES THE RELATIVE PRICE.

The difference in the flexibility results (26.8) and (26.9) is big in comparison with the limits of significance. The whole bunch-map analysis tells us beyond doubt that the money flexibility determined via butter is significantly larger than that determined via meat. But from theory the two determinations ought to be equal. The conslusion is therefore inescapable that there must be some factor which theory did not take account of, but which has been present in the data and distorted the results.

It will be remembered that the consumption surface theory assumed no substitution possibilities between the reference commodity and other commodities. The following questions therefore naturally present themselves: Does meat stand in substitution relation to other goods? Does butter stand in such relation? And can such relations explain the difference in flexibility measurement obtained?

Let us see how the consumption surface theory can be generalised to take account of substitution possibilities. Let x and y be the quantities consumed of two commodities, and let u and v be their marginal utilities. If there is no substitution possibility between the two commodities, u would be a function only of x, and v a function only of y. If $u(x)$ and $v(y)$ denote these functions, and p and q are the prices of the two commodities, we would have in the equilibrium of the market

$$(27.1) \qquad \frac{u(x)}{p} = \frac{v(y)}{q} = \frac{w(r)}{P}.$$

Leaving out the middle term of this equation, we get of course (26.1). In other words if u depends only on x and v only on y, we just have the case of two independent reference commodities which was the assumption underlying the preceding analysis of meat and butter.

If there exist substitution possibilities, u must be looked upon as a function both of x and y, and so must v. The equilibrium equation now takes on the form

$$(27.2) \qquad \frac{u(x,\,y)}{p} = \frac{v(x,\,y)}{q} = \frac{w(r)}{P}.$$

Consider for a moment the first of these equations. It defines y as a function of x and of the price ratio

$$(27.3) \qquad \lambda = \frac{p}{q}.$$

Let this function be

$$(27.4) \qquad y = f(x,\,\lambda).$$

Inserting the expression (27.4) for y in the first member of (27.2) we see that the equation of the consumption surface now takes the form

$$(27.5) \qquad w(r) = \alpha u(x,\,f(x,\,\lambda)).$$

The function in the right member of this equation is some function of x and λ which for brevity we denote

$$(27.6) \qquad U(x,\,\lambda) = u(x,\,f(x,\,\lambda)).$$

With this notation the equation of the surface of consumption takes the form

$$(27.7) \qquad w(r) = \alpha \cdot U(x,\,\lambda).$$

Thus the new element that comes in when substitution possibilities exist is that the function in the right member ρf (26.1) must be conceived of as depending not only on x but also on the ratio of the price of the reference commodity to that of the

other commodity with which it is in substitution relation. The isoquant principle must be generalised into a principle of constant quantity consumed *and* constant price ratio to the substitution commodity, but with this generalisation the principle applies.

If there are a number of commodities Nos. 1, 2...*n* with which the reference commodity is in substitution relation, we need to consider all the equilibrium equations

(27. 8)
$$\frac{u(x, y_1 y_2 \ldots)}{p} = \frac{v_1(x, y_1 y_2 \ldots)}{q_1} = \frac{v_2(x, y_1 y_2 \ldots)}{q_2} = \cdots = \frac{w(r)}{P}.$$

where $y_1 y_2 \ldots$, $q_1 q_2 \ldots$, $v_1 v_2 \ldots$ are the quantities consumed, the prices and the marginal utilities of those commodities with which the reference commodity is in substitution relation. Combining the various members in (27.8) with the first member, we get a system of equations which we may assume to be solved in the form

(27. 9) $y_i = f_i(x, \lambda_1 \lambda_2 \ldots)$ $(i = 1, 2 \ldots)$

where

(27. 10) $$\lambda_i = \frac{p}{q_i}.$$

In this general case the equation of the surface of consumption would be

(27. 11) $w(r) = \alpha \cdot U(x, \lambda_1 \lambda_2 \ldots)$

where

(27. 12) $U(x, \lambda_1 \lambda_2 \ldots) = u[x, f_1(x, \lambda_1 \lambda_2 \ldots), f_2(x, \lambda_1 \lambda_2 \ldots) \ldots]$

The formula (27.11) is a perfectly general formulation of the surface of consumption idea, and if the necessary data were available, the money flexibility could be determined as

(27. 13) $$\check{w} = \frac{d \log \alpha}{d \log r} = \frac{d\alpha}{dr} \cdot \frac{r}{\alpha}$$

along *any* curve on the empirically determined surface (27.11), where x, $\lambda_1 \lambda_2 \ldots$ etc. are all constants. Assuming for sim-

plicity the relations to be linear one may attempt a linear regression analysis including the $\lambda's$ as variates.

It is probable, however, that if an analysis was made according to the above perfect general scheme, it would be found that multiply collinear set would frequently occur.

As an example consider first the extreme case where the whole structure of the market is such that the prices of the various commodities with which the reference commodity is in substitution relations move (nearly) proportionally to the price of the reference commodity. In this case the function U in the right member of (27. 11) would (nearly) depend only on x. Since the isoquant method is independent of the shape of the function $u(x)$ (or $U(x)$) we see that in this case the correct money flexibility would be obtained by *applying the whole isoquant technique exactly as if there had been no substitution possibilities.* Thus it is only to the extent that the prices of the substitution commodities vary *disproportionally* to that of the reference commodity, that the effect of the substitution needs to be taken account of in the isoquant method of measuring money flexibility.

But, of course, if the same data are used to determine the elasticity of the reference commodity itself, one would *not* get the correct result by proceeding *as* if no substitution possibilities exist. Indeed, the elasticity of the function (27. 12) with respect so x, (under constant $\lambda's$) is not the same as the partial elasticity of $u(x, y_1 y_2 \ldots)$ with respect to its first variate.

The case of proportional price movements is only a very special case of multidimensional connections between the variates in the right member of (27. 11). Even without assuming any such proportionality it seems plausible — and indeed necessary — to take account of the fact that some sort of *supply relations* exist for the commodities between which substitution possibilities exist. By taking account of this we shall not only protect ourselves against some of the risks of falling into multicollinear situations, but we shall also obtain a theoretical scheme which is much simpler and more amendable to statistical analysis.

Let us assume that for each of the substitution commodities there exists — besides the equations (27.9) for this commodity — some other relation that connects quantities and prices. Let this system of relations be

(27.14) $S_i(x, p, y_1 y_2 \ldots, q_1 q_2 \ldots) = 0$ $(i = 1, 2 \ldots)$

From this system and (27.9) we may imagine that both the $y's$ and the $q's$ are expressed in terms only of x and p

(27.15) $y_i = g_i(x, p)$ $q_i = G_i(x, p)$ $(i = 1, 2 ..)$

Inserting this into (27.8) we see that the equation of the surface of consumption now takes on the form

(27.16) $w(r) = \alpha \cdot U(x, p)$

where U is a function only of the two variates x and p, namely

(27.17) $U(x, p) = u(x, g_1(x, p), g_2(x, p) \ldots)$

In other words, the only new variate which we now need to take account of in the right member of the equation of the surface of consumption is the absolut price of the reference commodity. The absolute price is, so to speak, taken as a catch-all for all the various possible substitution effects that may concern this commodity in its capacity of reference commodity for money flexibility measurements. We do not go into any specification of *which* other commodities the reference commodity is in substitution connections with.

The assumption (27.16) was used on the meat and butter data. The relation was assumed linear in log p, so that the only modification in the preceding analysis consisted in including the variate

(27.18)
$$x_7^{(\text{meat})} = \log p^{(\text{meat})} - 1.4$$
$$x_7^{(\text{butter})} = \log p^{(\text{butter})} - 1.5$$
(1.4 and 1.5 being provisoric means).

Since concentric numbering of the subsets had been used, the additional tilling work needed could be performed simply by elongating the lists and tables previously used. The result of the computations are represented graphically in Figures 13 and 14 in the preceding Section. We shall now consider these charts in their entirety, not only the cells belonging to the first six variates. We shall go through the analysis anew, this time taking into account also those subsets where 7 occur.

We shall take meat first. The three-set (126) in meat is still

the only one that can be used, and in the four-sets we find as before that only those where 6 occur has any organisation. This leaves us with only one new four-set to be considered, namely (1267). This set is quite remarkable. The organisation here is very good. Indeed, it is markedly better than the organisation in the best four-sets we had before 7 was added. And the change in slope is quite conspicuous. We notice that there is a real fight between 5 and 7 in the influence on the slope. The inclusion of 5 (the trend factor) had — as we noticed it in the preceding analysis — a tendency to increase the flexibility, while we now see that the inclusion of 7 (the substitution factor) tends to *lower* the flexibility i. e. to flatten the slope: in the set (1267) the flexibility is 0.75 as against 1.03 in the set (1256). This suggests that substitution does play a role, and that it must be taken into account in order to get a correct measurement of the money flexibility. Of course the variate x_7 (price) now considered does not represent the ordinary demand curve connection between price and quantity, that is already taken into account by the variate α; x_7 represents a *deviation* from the regular demand curve, namely a deviation caused by substitution.

The difference in the slopes in the various four-sets that appear significant makes it necessary to consider the five-sets. Amongst the five-sets containing 7, (12467) shows an excellent tightness. It is suggestive of the importance of 6 to compare (12457) with (12467). Further, we see that the fight between 5 and 7 is also quite marked in the five-sets. In the two admissible sets where 5 (and not 7) is present, namely (12356) and (12456) the flexibility is 1.11 and 1.06 respectively, while in the two admissible sets where 7 (but not 5) is present, namely (12367) and (12467) the flexibility is 0.80 and 0.75 respectively. If the difference between these slopes is a gross-slope-effect, it seems probable that the true flexibility must be somewhere between these values, say between 0.90 and 0.95. It is interesting to compare this with the result in the set (12567) where *both* 5 and 7 are present (besides (126) which always prove to be necessary for a good fit). The tightness of (12567) is none too good, but it does show some organisation, and the flexibility measured by the diagonal regression is in the range 0.90 to 0.95.

Also in the six-sets is the fight between 5 and 7 the domi-

nating feature; compare, for instance, (123456) which contains 5 (trend) but not 7 (substitution) and which gives the flexibility of 1.11, with (123467) which contains 7 but not 5 and gives the flexibility of 0.79. If both trend and substitution is taken into account, it seems plausible that the flexibility will turn out to be somewhere midway between these figures, say 0.95. This also checks fairly well with (124567) which contains both 5 and 7, and gives the flexibility 0.99. The seven-set is of interest because it has sufficient tightness to indicate that there is no "explosion" and that the slope difference observed in the six-sets is consequently a gross-slope-effect so that it has a meaning to take some sort of an average between them. But the slope in the seven-set is not so clearly defined that it seems advisable to adopt the diagonal flexibility in this set, as a better result than the above average of 0.90 to 0.95. Taking into account all available facts, it seems that when the substitution possibilities are eliminated, we may put

$$(27.19) \qquad - \check{n}^{(\text{via meat})} = \text{between } 0.95 \text{ and } 1.00$$

with a range of variation of some 5 per cent up or down.

In conclusion we may thus say that substitution has some effect — but not a very great one — in the case of meat. The money flexibility measured via meat is *lowered* when this correction is taken into account.

Now, as to butter. Here there is a high gross correlation between 1 and 7, namely $r_{12} = -0.954$. This is simply due to the fact that the butter price p has fluctuated much more violently than the cost of living price P. When this is the case, there will of course be produced a high (negative) correlation between the variates 1 which is $(\log P - \log p)$, and 7 which is $\log p - 1.5$. We must therefore be prepared to find that the inclusion of the variate 7 with those already discussed will in the case of butter produce a much larger *opening* of the bunches, i. e. a less definite slope indication than in the previously considered sets not including 7. That this actually happens is seen by a glance at the bunch map in Figure 2 of Section 26. The question is whether the change in slope produced by the inclusion of 7 is so marked that significant conclusions can be drawn even though the new bunches are more open. The whole situation is such as to put our apparatus of confluence criteria to an interesting test.

The new three-set obtained by adding 7, namely (127) does not show a very good tightness, but at least so much of it that the set cannot be disregarded altogether. The explanation of the amount of organisation here found is probably as follows. The variate 1 is ($\log P - \log p$) and 2 is ($\log \varrho - \log P$), ϱ being the nominal income; since $\log P$ occurs here in both expressions and with opposite sign, some negative correlation between 1 and 2 may be expected. There actually turns out to be very little of it (as is seen by the appearance of the bunch in the set (12)). This is due to the variation of $\log p$ and $\log \varrho$. But if the effect of $\log p$ is eliminated, as it is in the set (127), the fit is improved. The (12) slope is not — 1 as it would have been if it had expressed only the definitional connection due to the terms $\log P$ and — $\log P$. The slope is flatter, which is explained by the presence of $\log \varrho$ in the variate No. 2. Indeed on account of the whole economic structure, ϱ and P tend to vary in the *same* direction, and this will work counter to the negative connection between 1 and 2 that is created by the terms $\log P$ and — $\log P$. The question of whether this may be responsible for a "spurious" element in the determination of the money flexibility will be discussed in Section 28. At present we shall continue the systematic analysis of the bunch map.

In the four-sets we notice the same fight between 5 and 7 as we found in meat: 5 steepens the slope (a little) and 7 flattens it (quite a bit). It is true that the organisation of the four-bunch containing 7 (in addition to the fundamental set (126)) is not good (the reasons for this we have already discussed). But even though the bunch in (1267) is rather open, it is quite obvious that it represents a flatter slope than in (1256). In (1267) the diagonal flexibility is 0.80 as against 1.47 in (1256). The big difference between these two measurements makes it necessary to consider the five sets.

The five-set that contain 5 and 7 (in addition to (126)) is markedly tighter than (1267), and the steepening effect is clear: 5 is therefore useful when added to (1267). In going from (1256) to (12567) the change is less clear. There is — as in all cases where 7 is added to a set containing (126) — a marked opening of the bunch. The change in slope is somewhat veiled by the smaller tightness in the new bunch, but some flattening seems to be present. The diagonal flexibility in (12567) is 1.26 as against 1.47 in (1256).

After the inclusion of 7 there is also another feature which becomes apparent. 4, and to a still higher degree, 3 have a tendency to flatten the slope as compared with 5. This effect is very marked: for instance, if 3 is added to (1267) we get a diagonal flexibility of 0.63, while if 5 is added, we get 1.26. And the bunches, while not very tight, still have so much organisation that they cannot be disregarded. All this makes it necessary to consider the six-sets. The tightness here is not any poorer than in the five-sets, and the pulls of the various variates 3, 4, 5 and 7 in different directions are still manifest. This suggests that as a final attempt we must go to the big seven-set. This final step turns out to be rather significant. The tightness in this bunch is decidedly better than in any of the six-sets with the exception of (123456). And with regard to the passage from this latter set the change in slope is here very conspicuous. One would therefore, without hesitation, characterise all the variates in the seven-set as useful. The regression in this set must consequently be considered as admissible. In this regression the diagonal flexibility is

$$(27.20) \qquad - \check{w}^{(\text{via butter})}_{(1234567)} = 0.97 \cdot / \cdot 0.75.$$

Thus by taking into account substitution possibilities, we get sensibly the same results for the money flexibility via meat and via butter. Not only that, but the values found are now closer to that which I found by using the United States household budget of 1918/19. Still there is a good bit of difference, the average values on the flexibility curve I found ranged around 0.5. The difference may in part be due to the fact that I was using exclusively city families, while the present data include the United States as a whole. Another, and probably more important, fact is, I think, what may perhaps be called the "silk-shirt mentality". The data for my investigations were collected in a period of prosperity when incomes were increasing and people were in an "active" consumption mood: they wanted to expand consumption both quantitatively and to new categories of goods. In other words, they were just in that situation which would give a low money flexibility. The data used in the present study covers a period where there have been both ups and downs, and certainly not any steady "silk-shirt mentality".

28. DEFINITIONAL FLATTENING AND SPURIOUS CORRELATION.

In Section 7 we considered each observational variate as a linear form in certain basic — not directly observable — variates. What further complications may arise if some of the observational variates can be expressed as linear forms in some of the other *observational* variates? To what extent will this cause "spurious" correlation? I have preferred not to discuss this problem in the theoretical part III, but rather to treat it in connection with the study of the consumption data: this will make the discussion more concrete.

Consider the two variates

$$(28.1) \qquad \begin{aligned} x_1 &= \log P - \log p \\ x_2 &= \log \varrho - \log P \end{aligned}$$

that occur in the flexibility study, p being the price of the reference commodity, P the cost of living and ϱ the nominal income. The fact that x_1 contains $\log P$ and x_2 contains $- \log P$ will tend to produce a negative correlation between x_1 and x_2. Will this cause "spurious" results? That will depend on the nature of the variability of $\log p$, $\log P$ and $\log \varrho$. If there is no significant correlation between these latter variates, then the observed correlation between x_1 and x_2 would have to be interpreted as spurious. But if there is some significant connection between all or some of the variates $\log p$, $\log P$ and $\log \varrho$ — as there certainly is, for instance between $\log P$, (P being the cost of living), and $\log \varrho$, (ϱ being the nominal income), these two variates having roughly speaking a tendency to move in the same direction — then an observed correlation between x_1 and x_2 cannot be interpreted only as spurious. It may *in part* be spurious, namely to the extent that $\log p$, $\log P$ and $\log \varrho$ contain *outside* components, this is to say, components that must be looked upon as disturbances in the system of the other variates considered. But to the extent that the observed correlation between x_1 and x_2 is due to the *systematic* components of $\log p$, $\log P$ and $\log \varrho$, we cannot consider it as spurious.

If we did that, we could just as well turn the matter around and say for instance: We have

$$(28.2) \qquad \begin{aligned} \log P &= x_1 + \log p \\ \log \varrho &= x_1 + x_2 + \log p \end{aligned}$$

and since the observable variate x_1 here occurs both in the expressions for log P and log ϱ we must be prepared to find a "spurious" correlation when we observe log P and log ϱ together. It is clear that this argument leads in a circle; indeed, whatever variates we have given, we can always introduce some new variates linearly dependent on the old, and then write some of the old variates as linear forms where some of the new variates occur. The mere fact that two observational variates can be expressed as linear forms in certain other observational variates does therefore not in itself constitute a "spurious" element. All depends on the nature of the variability of these other variates by which the first variates are expressed.

In particular if we have arrived a priori at certain theoretical relations whose coefficients we attempt to determine statistically, it may well be that some of the observational variates entering can be expressed in terms of others. This transformation may then only be an *alternative* way of formulating the problem and need not introduce anything "spurious". A successful determination of the coefficients sought may be possible by using either form of the equations. Wether the attempt shall meet with success, will depend primarily on the actual strength with which the postulated *structural* relation exists in the variates. If this underlying structural relation is strong, it will be reflected and may be measured in either form of the equations.

One thing we must look out for, however, is that a number of degrees of freedom is left which is sufficient for the regression analysis contemplated. By this I mean the following. Quite generally let

(28. 3) $$x_1 x_2 \ldots x_L$$

be a set of variates which are or could be observed.

In the above flexibility example we could, for instance, consider x_1 and x_2, and furter log p, log P, log ϱ and log x, x being the quantity consumed, giving a total of $L = 6$ variates. All of these would of course not be independent, (28. 1) (or if we prefer (28. 2)) constitute two independent equations between the six variates. Therefore, if we plotted a scatter diagram in 6-dimensions, this scatter would have an unfolding capacity of not more than 4. We shall say that this scatter

has a *definitional* flattening of 2. In addition to this some *structural* flattening may exist, for instance, the one represented by the surface of consumption. Quite generally, if between those variates that are included in an observed scatter there exists by definition φ independent linear equations, we shall say that the scatter has a definitional flattening of φ. If in addition the structural conditions impose ψ further relations between the variates, we shall say that the *structural* flattening is ψ. The number $\varkappa = \varphi + \psi$ is the *total* flattening. And it is this total flattening that is revealed by the confluence analysis.

In order that a regression equation fitted to the data shall give any information at all about a *structural* relation present in the data, it is of course necessary that the scatter to which the equation is fitted has no definitional flattening: indeed, a regression equation has a sense only when the scatter is flattened exactly once, and if exactly one definitional flattening is present, this is the only thing that will show up in the empirical regression equation fitted to the data. A definitional flattening is indeed always mathematically exact, while the structural flattenings are always somewhat blurred by the presence of the disturbances. They cannot therefore compete with the definitional flattening in influencing the result obtained. If there are *more* than one definitional flattening in the set of those variates that are subjected to the regression analysis, the regression coefficients will of course be exactly of the indeterminate $\frac{0}{0}$ form.

As examples we may notice that there is no definitional flattening in the set consisting of the two variates $x_1 x_2$ considered above, nor is there any such flattening if $\log x$ ($x =$ quantity consumed) is added to x_1 and x_2, indeed even taking into account the way in which x_1 and x_2 are defined (in terms of p, P and ϱ) x_1 and x_2 and $\log x$ may be chosen quite arbitrarily. Nor is there any definitional flattening in the set consisting of the three variates $x_1 x_2$ and $\log p$. But in the set of four variates x_1, x_2, $\log p$ and $\log \varrho$ there exists a definitional flattening, namely the last equation in (28. 2). A regression in this set of 4 variates would give a mathematically exact fit and would therefore be useless for any investigation of structural relations.

None of the sets actually used for flexibility measurements in the preceeding Sections contains a definitional flattening, but

conceivably the variability nature of the basic variables involved may be such that some spurious effect is produced. For instance, if it should happen that log p, log P, log ϱ and log x were entirely disorganised without being influenced by that systematic connection which we are looking for, namely the surface of consumption, then the observed correlation in the various x_i and in particular the (12) slope would be spurious, caused by the fact that the x_i are linear forms in the variates log p, log P. etc. In particular the fact that the (12) slope happens to turn our rather close to — 1, that is to say equal to that value which in view of (28.1) would come out if log p, log P and log ϱ were disorganised, may make us suspicious.

However, there are various reasons for rejecting this alternative. In the first place it is both for theoretical and empirical reasons out of the question to assume log p, log P and log ϱ independent. Roughly speaking there will be a tendency for P and p to move together, and also a tendency for ϱ and P to move together. From (28.1) it is therefore seen that what the (12) coefficient does measure is the amount of *deviation* which the material shows from the proportionality between P and p on the one hand and ϱ and P on the other. This deviation can of course go in either direction. When we have found that it does go so definitely in the negative direction (giving a negative (12) slope), this must be due to some structural relation in the data. Furthermore, we have in all instances noticed that we get organisation in the bunches only by taking into account the quantity consumed, which is a variate that has no definitional connection whatsoever with p, P or ϱ. This is another strong indication that the result expresses a structural connection, not simply a spurious relation due to the presence of log P in x_1 and — log P in x_2.

Finally, we may test the spurious factor in the (12) coefficient by the following experimental modification of the data. Let x_1 and x_2 be defined by (28.1) and let us take exactly those values of p, P and ϱ that were observed in the United States butter data 1919—31 and used in the analysis of Sections 26 and 27. But let us imagine that the quantity consumed had been different. Let us put its logarithme equal to

$$(28.4) \qquad x_8 = x_1 - 0.7\,x_2 + 0.12\,x_7 + e$$

where x_1 and x_2 are defined as above, x_7 is log p (as in Section

(27) and e an eratic variate (lottery drawings) whose standard deviation is about 15 per cent of that of x_8. The equation (28.4) would now be the modified form of the surface of consumption. The term with x_7 in (28.4) represents the substitution effect. We have chosen the coefficient in front of x_2 *negative;* this means that the money flexibility will now be *positive,* namely $+0.7$. Of course, this is a very unrealistic assumption, and it is made here only to get a clear cut case where the money flexibility is definitely different from -1. The object of the experiment is just to see if, by using the same method as in the preceding Sections (now with x_8 instead of x_6) we obtain that value for the money flexibility which would now be correct, namely $+0.7$ or if we still obtain some value near -1. The result of the computations are given in Figure 15. (The scale in 1278 is 10 times the scale in 128).
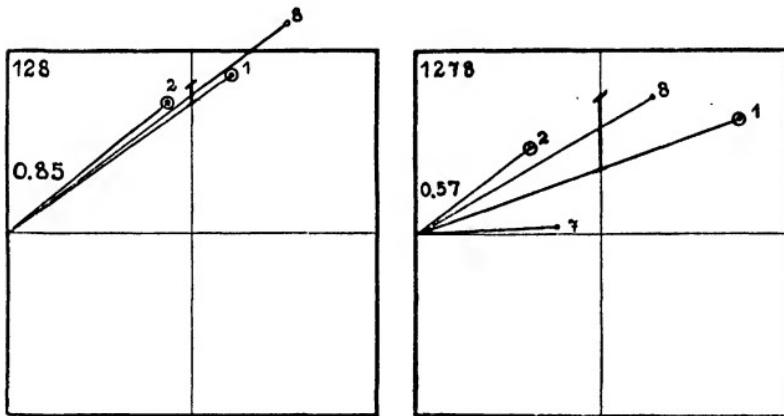
Fig. 15.

From these graphs we see immediately that the general slope of the (12) connection is now *positive.* Further, we notice that the inclusion of 7 to the set that from theory is the fundamental one, namely (128) will also now make the fit poorer. This is what we would expect since there exists such a high correlation between 1 and 7. (Both 1 and 7 are, it will be remembered, the actual data). The opening of the bunch is, however, not so strong as completely to veil the fact that some change in slope is produced. By the standard rules for the analysis of the bunch map it appears admissible to consider the four-set (1278). In this set the diagonal flexibility turns out to be

$$(28. 5) \qquad + \check{w}^{(\text{via constr.})}_{(1278)} = 0.57 \ \cdot / \cdot \ 0.69.$$

It is seen that the "true" flexibility according to (28. 4), namely $+0.7$ is amply contained between the upper and lower limits of (28. 5), namely $+0.83$ and $+0.39$. The empirical value 0.57 is not so very much off the true value, the discrepancy that does exist is of course due to the rather heavy erratic component which we have purposely introduced in x_8.

Since the money flexibility in the determinations of Section 27 were decidedly negative, and since all the data are now exactly the same as before, except the variate which is now denoted as No. 8 and made to fit the artificial surface of consumption (28. 4), we must conclude that this surface when it exists does play a dominating role in determining the observed flexibility. It seems that this flexibility as determined by the actual data is not due to a spurious connection between x_1 and x_2 created through the presence of log P and $-$ log P in these variates, but is a reality.

29. SCATTERANCES AND LINE COEFFICIENTS IN THE CONSUMPTION DATA.

The scatterances in the meat and butter data are given in tables (29. 1) and (29. 2) respectively.

In tables (29. 1) and (29. 2) are indicated also the sums of the scatterances on each level. These sums are nothing but the coefficients A_k of the characteristic polynomial $P(\lambda)$. These coefficients are used amongst other for checking purposes in connection with (29. 4).

Consider meat first. None of the two-sets show any good correlation. If the data are to have any meaning in flexibility analysis, we must therefore proceed to the three-sets. This is also in agreement with the theoretical pattern which requires the inclusion of at least (126).

Amongst the three-sets (126) stand out very conspicuously. The scatterance in this set is equal to 0.049, while the smallest of the other three-rowed scatterances — that in (157) — is 0.136. The subscatterances in (126), namely 0.69, 0.73 and 0.83 are not widely different, but they are so large that it seems safe to consider (126) as a promising set according to (III) of Section 1. If a three-set is to be accepted, it must undoubtedly be (126).

TABLE (29.1). SCATTERANCES, MEAT 1919—31.

Two-sets	Δ	Three-sets	Δ	Four-sets	Δ	Five-sets	Δ
12	0.698223	123	0.261244	1234	0.088678	12345	0.029444
13	0.955034	124	0.260963	1235	0.106864	12346	0.005990
23	0.432474	134	0.396221	1245	0.086653	12356	0.004493
14	0.822088	234	0.303500	1345	0.192124	12456	0.003716
24	0.892555	125	0.290416	2345	0.100775	13456	0.051034
34	0.714656	135	0.477426	1236	0.018238	23456	0 033403
15	0.558982	235	0.213892	1246	0.018074	12347	0.041601
25	0.536446	145	0.453729	1346	0.111493	12357	0.022068
35	0.856850	245	0.296388	2346	0.108935	12457	0.017472
45	0.958412	345	0.447485	1256	0.012920	13457	0.045917
16	0.738746	126	0.049302	1356	0.166484	23457	0.047471
26	0.838212	136	0.338659	2356	0.108264	12367	0.006865
36	0.752682	236	0.324755	1456	0.136980	12467	0.006671
46	0.383172	146	0.282653	2456	0.098441	13467	0.053184
56	0.981904	246	0.321039	3456	0.167140	23467	0.068580
17	0.503327	346	0.267787	1237	0.128173	12567	0.002771
27	0.912882	156	0.370955	1247	0.123609	13567	0.035001
37	0.997420	256	0.282881	1347	0.190485	23567	0.027692
47	0.915905	356	0.540812	2347	0.224222	14567	0.028330
57	0.992350	456	0.366580	1257	0.062455	24567	0.023410
67	0.870223	127	0.342973	1357	0.114387	34567	0.128421
$A_2 = 16.312543$		137	0.470964	2357	0.168659	$A_5 = 0.683534$	
		237	0.365366	1457	0.108869		
		147	0.413727	2457	0.141342	Six-sets	Δ
		247	0.665229	3457	0.374405		
		347	0.612243	1267	0.018567	123456	0.001216
		157	0.136530	1367	0.163270	123457	0.005820
		257	0.476831	2367	0.204509	123467	0.002209
		357	0.849981	1467	0.142228	123567	0.000928
		457	0.877011	2467	0.205057	124567	0.000749
		167	0.371832	3467	0.212269	134567	0.010495
		267	0.535779	1567	0.079731	234567	0.007620
		367	0.602123	2567	0.079977	$A_6 = 0.029037$	
		467	0.333395	3567	0.425548	Seven-set	
		567	0.852954	4567	0.318429		
		$A_3 = 14.453625$		$A_4 = 5.008254$		1234567	0.000240

TABLE (29. 2). SCATTERANCES, BUTTER 1919—31.

Two-sets	Δ	Three-sets	Δ	Four-sets	Δ	Five-sets	Δ
12	0.994990	123	0.507082	1234	0.233744	12345	0.013862
13	0.687733	124	0.850195	1235	0.076096	12346	0.012990
23	0.787931	134	0.318713	1245	0.055939	12356	0.002742
14	0.952541	234	0.399923	1345	0.167600	12456	0.002010
24	0.892555	125	0.177620	2345	0.106256	13456	0.027658
34	0.507588	135	0.446881	1236	0.031726	23456	0.017060
15	0.699055	235	0.232571	1246	0.047579	12347	0.007747
25	0.536446	145	0.658753	1346	0.153278	12357	0.003087
35	0.984998	245	0.296399	2346	0.110746	12457	0.003496
45	0.958417	345	0.486056	1256	0.006566	13457	0.005524
16	0.736198	126	0.073045	1356	0.108277	23457	0.019409
26	0.398256	136	0.331648	2356	0.063126	12367	0.001633
36	0.982863	236	0.262579	1456	0.108763	12467	0.003212
46	0.929687	146	0.559087	2456	0.055070	13467	0.005281
56	0.378311	246	0.355347	3456	0.086720	23467	0.021658
17	0.090008	346	0.468855	1237	0.029247	12567	0.000411
27	0.994272	156	0.257888	1247	0.057532	13567	0.004552
37	0.843822	256	0.145874	1347	0.011595	23567	0.004519
47	0.959484	356	0.320887	2347	0.299388	14567	0.006851
57	0.565292	456	0.181146	1257	0.011280	24567	0.004150
67	0 641913	127	0.069050	1357	0.021117	34567	0.034049
$A_2 = 15.522360$		137	0.042893	2357	0.068889	$A_5 = 0.201901$	
		237	0.598477	1457	0.041493		
		147	0.085693	2457	0.054938	Six-sets	Δ
		247	0.836323	3457	0.204048		
		347	0 422533	1267	0.004999	123456	0.000435
		157	0.044422	1367	0.020266	123457	0.000454
		257	0.163961	2367	0.062976	123467	0.000420
		357	0.457940	1467	0.039623	123567	0.000112
		457	0.537318	2467	0.070323	124567	0.000122
		167	0.054503	3467	0.192534	134567	0.000910
		267	0.104708	1567	0.016320	234567	0.001188
		367	0.406681	2567	0.012452	$A_6 = 0.003641$	
		467	0.467206	3567	0.126678	Seven-set	Δ
		567	0.207688	4567	0.090327		
		$A_3 = 11.829945$		$A_4 = 2.847511$		1234567	0.000015

The four-sets with the lowest scatterances are summarised in (29.3).

TABLE (29.3). LOWEST FOUR-SETS, MEAT.

	1256	1246	1236	1267
Scatterance	.0129	.0180	.0182	.0185
Subscatterances	.290	.260	.261	.049
	.049	.049	.049	.342
	.370	.282	.338	.371
	.282	.321	.324	.535

These four-sets form a group quite distinct from the other four-sets, they have markedly lower scatterances. The lowest four-rowed scatterance after those listed in (29.3) is 0.062 in (1257). It will be noticed that the four-sets listed in (29.3) are those obtained by adding 3, 4, 5 and 7 respectively to (126). Since (126) had markedly lower scatterance than the other three-sets, it was to be expected that the four-sets containing (126) would give low scatterances. The interesting thing is that actually no other four-set comes in and competes with those that derive their good fit from the connection with (126). We also see that all these four-sets have a marked spread in their subscatterances. All the four-sets listed in (29.3) may therefore be considered as promising. This checks with the bunch map analysis where it was found that just these four-sets were the ones to be considered.

In the five-sets a similar effect is found: the 6 sets obtained by adding any two of the variates 3, 4, 5 and 7 to (126) form again a group by themselves having scatterances lower than the others. Again, it is interesting to note that no other set comes in. There is also sufficient spread in the subscatterances to consider any of these sets promising. This also checks with the bunch analysis.

But when it comes to *discriminating* between the 6 promising five-sets, the scatterance analysis proves inadequate. We then need to take into account finer traits of the data which are not revealed by the scatterances. From this point on we must leave the scatterance and rely on the bunch-map analysis. This is a good example of the general rule that the scatterances may be useful in indicating roughly the simpler features of the data — particularly in cases where these features are very distinctly and strongly present. But the scatterances must not be pressed to give information about the finer features: they will then only give nonsensical information.

The most conspicuous difference between butter and meat is that in butter we find already a two-set with a small scatterance, namely (17). The reason for this was discussed in Section 28. It is mainly a spurious correlation and should not be taken as an expression for a structural relationship. The second lowest two-set in butter is (56) with a scatterance of 0.378. This simply expresses the fact that there is a strong *trend* in the butter consumption. The third lowest is (26) with a scatterance of 0.398. More precisely the variates 2 and 6 move in the same direction, r_{26} being positive. This can probably be interpreted as a structural relation, namely the *Engel curve* for butter: Butter consumption increases as income increases. Of course, the connection as exhibited in the gross correlation between these two variates is not perfect, due to the influence of other factors. The main conclusions from the study of the three and higher rowed scatterances for butter are similar to those for meat; the reader can himself easily carry this analysis through.

Table (29. 4) gives the line coefficients for butter. The two-rowed line coefficients are the same as the scatterances and need therefore no further comment.

The three-rowed line coefficients indicate (126) as the best set. This is remarkable. The line coefficient in this set is even smaller than are that in (127) and in the other sets containing the spuriously connected variates 1 and 7. Amongst the four-sets the best is (1256) with a line coefficient of 0.16. Its sub-line coefficients are 0.33, 0.15, 0.73 and 0.77. These show a considerable spread and some are not small: (1256) is therefore a decidedly promising set. This checks with the bunch analysis which, it will be remembered, gave best tightness in (1256).

If we want to go from the four-sets to the five-sets, we see from table (29.4) that we must accept a marked increase in the line coefficient. Indeed, the lowest five-rowed line coefficients range about 0.25 as against 0.16 in the four-sets. This means that even the best five-set is such that the *average* tightness in its various bunches is poorer than in the set (1256). Whether it is nevertheless admissible to consider some of these sets will depend on the strength of the slope changes. We have therefore here reached a point where the analysis cannot be pushed any further, purely on the basis of line coefficients. From this point on we have to rely on the bunch map.

TABLE (29. 4). LINE COEFFICIENTS, BUTTER.

Three-sets		Four-sets		Five-sets	
123	0.75678	1234	0.73496	12345	0.39597
124	0.94956	1235	0.54154	12346	0.44792
134	0.64270	1245	0.21205	12356	0.25537
234	0.77093	1345	0.59956	12456	0.26105
125	0.33214	2345	0.67764	13456	0.72467
135	0.72594	1236	0.31049	23456	0.76808
235	0.40395	1246	0.21117	12347	0.25577
145	0.87881	1346	0.70943	12357	0.55064
245	0.48107	2346	0.60937	12457	0.69146
345	0.75535	1256	0.16191	13457	0.30516
126	0.15412	1356	0 71003	23457	0.44367
136	0.52205	2356	0.75774	12367	0.50031
236	0.53238	1456	0.53301	12467	0.61122
146	0.73847	2456	0.65367	13467	0.29923
246	0.68533	3456	0.53726	23467	0.47551
346	0.74637	1237	0.29905	12567	0.52052
156	0.73153	1247	0.28732	13567	0.41245
256	0.77147	1347	0.20054	23567	0.31695
356	0.56212	2347	0.80703	14567	0.64992
456	0.33853	1257	0.61687	24567	0.44963
127	0.17730	1357	0.28626	34567	0.70898
137	0.17918	2357	0.44407	Six-sets	
237	0.78496	1457	0 34145		
147	0.23671	2457	0 23847	123456	0.28246
247	0.93000	3457	0.72540	123457	0.61304
347	0.73485	1267	0.57753	123467	0.56814
157	0.26183	1367	0.38595	123567	0.64833
257	0.35168	2367	0.45742	124567	0.52312
357	0.73664	1467	0.41675	134567	0.50627
457	0.80447	2467	0.32958	234567	0.48650
167	0.27406	3467	0.80264	Seven-set	
267	0.24353	1567	0.49755		
367	0.60943	2567	0.30720	1234567	0.60570
467	0.66595	3567	0 78466		
567	0.76023	4567	0.58951		

The line coefficients in the two-sets are simply the scatterances given in (29. 2).

30. EMPIRICAL DISTRIBUTION OF SCATTERANCES IN AN EIGHT-SET OF RANDOM VARIATES.

For various purposes it is useful — as a standard of comparison — to know the decline that takes place in the scatterances as we pass to higher sets in the case where no systematic

connections exist between the variates. I have therefore had computed the scatterances in all possible subsets of 3 eight-rowed correlation matrices formed on the basis of random data. The variates were constructed by lottery drawings. The number of observations from which the correlations were formed was $N = 100$. As an example the cumulated frequency distribution of the four-rowed scatterances contained in one of these matrices (No. III) is given in Figure 16. The complete list of the above mentioned scatterances computed with six decimal places are available at the University Institute of Economics, Oslo. A copy will be sent on request.

Fig. 16.

The median values and quartiles of the empirical distributions of these scatterances are given in tables (30. 1) — (30. 3).

TABLE (30. 1). MEDIAN OF THE EMPIRICAL DISTRIBUTION OF SCATTERANCES IN AN EIGHT-SET OF RANDOM VARIATES.

Set	Matrix No.		
	I	II	III
2-rowed	0.991	0.995	0.995
3 »	0.967	0.972	0.976
4 »	0.925	0 949	0.933
5 »	0.867	0.908	0.888
6 »	0.797	0.853	0.827
7 »	0.731	0.788	0.771
8 »	0.675	0.746	0.715

TABLE (30. 2).

Set	Lower quartiles in matrix No.			Upper quartiles in matrix No.		
	I	II	III	I	II	III
2-rowed	0.979	0.984	0.986	0.998	0.998	0.999
3 »	0.939	0.964	0.954	0.976	0.985	0.984
4 »	0.885	0.927	0.900	0.948	0.960	0.956
5 »	0.829	0.874	0.859	0.905	0 927	0.911
6 »	0.776	0.816	0.811	0.844	0.880	0.856
7 »	0.705	0.776	0.759	0.751	0.816	0.785
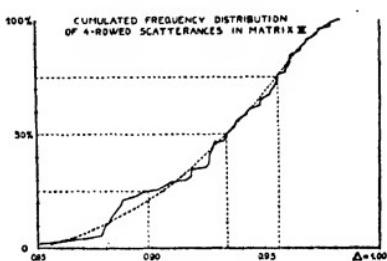
TABLE (30.3). AVERAGE OF CHARACTERISTICS FORMED IN ABOVE
3 MATRICES.

Set	Median	Lower quartile	Upper quartile
2-rowed	0.994	0.983	0.998
3 »	0.972	0.952	0.982
4 »	0.936	0.904	0.955
5 »	0.888	0.854	0.914
6 »	0.826	0.801	0.860
7 »	0.763	0.747	0.784
8 »	0.712		

Of course, these distributions do not represent the sampling
distribution of independent scatterances. For instance, some
of the four-rowed scatterances whose distribution is given in
Figure 16 are connected because they are all contained in one
and the same eight-rowed matrix. The thing that interests us
from the confluency view-point is this kind of distribution more
than the distribution of independent scatterances.

Some information about the distribution of independent
scatterances may however also be derived from the above data.
In Table (30.4) are listed the actual values of some scatteran-
ces that are not connected.

TABLE (30.4). INDEPENDENT SCATTERANCES OF RANDOM VARIATES.

Number of variates	Set	Matrix No.		
		I	II	III
2-rowed	12	.982556	.999261	.991714
	34	.980425	.981367	.999166
	56	.998239	.973878	.998607
	78	.997071	.981084	.997393
3-rowed	123	.975821	.915083	.937775
	456	.962942	.963738	.987513
4-rowed	1234	.949019	.887814	.932655
	5678	.895032	.926488	.942310
5-rowed	12345	.926138	.851324	.924423
6-rowed	123456	.871833	.816179	.889732
7-rowed	1234567	.705379	.775624	.811419
8-rowed	12345678	.675130	.745849	.714657

TABLE (30. 5). CHARACTERISTICS OF THE EMPIRICAL DISTRIBUTION OF SCATTERANCES OF RANDOM DATA. (For each characteristic several independent determinations were made by means of the computations in the 3 matrices I, II and III above. For instance each median given in the first column of the present table is an average of several determinations. Similarly for the quartiles.)

Dimensionality of the big matrix containing the scatterances whose distribution is sought	Dimensionality of the scatterances	Median value of the scatterances	Lower quartile	Upper quartile	Number of independent determinations of each characteristic	Lowest value found for the median	Highest value found for the median
2	2-rowed	.990	Same as median		12	.974	.999
3	2 »	.989	.973	.995	6	.965	.999
	3 »	.958	Same as median		6	.915	.995
4	2 »	.993	.977	.998	6	.986	.998
	3 »	.962	.942	.981	6	.944	.974
	4 »	.922	Same as median		6	.888	.949
5	2 »	.994	.966	.998	3	.992	.997
	3 »	.977	.948	.984	3	.970	.986
	4 »	.932	.992	.964	3	.888	.955
	5 »	.885	Same as median		3	.839	.927
6	2 »	.995	.985	.998	3	.993	.997
	3 »	.977	.962	.985	3	.970	.985
	4 »	.945	.918	.962	3	.933	.956
	5 »	.902	.881	.931	3	.863	.929
	6 »	.859	Same as median		3	.816	.890
7	2 »	.993	.982	.998	3	.991	.995
	3 »	.968	.951	.980	3	.962	.973
	4 »	.928	.900	.954	3	.907	.940
	5 »	.880	.848	.911	3	.834	.904
	6 »	.818	.802	.849	3	.778	.859
	7 »	.764	Same as median		3	.705	.811
8	2 »	.994	.983	.998	3	.991	.995
	3 »	.972	.952	.982	3	.967	.976
	4 »	.936	.904	.955	3	.925	.949
	5 »	.888	.854	.914	3	.867	.908
	6 »	.826	.801	.860	3	.797	.853
	7 »	.763	.747	.784	3	.731	.788
	8 »	.712	Same as median		3	.675	.746

The material that served to determine (30. 3) may of course also be utilised to find the corresponding medians and quartiles for the distribution of scatterances contained in a matrix of *lower* dimensionality than eight to which (30. 3) refers. Indeed, if we leave out all those scatterances that contain the affix No. 8 we obtain the corresponding distribution of scatterances contained in a seven-set. Similarly, one may leave out two affixes, and thus determine the distribution of scatterances contained in a six-set, etc. The result of these computations are given in table (30. 5).

31. AN EXAMPLE IN 8 VARIATES FROM THE NEW ENGLAND POTATO MARKET.

The examples of Sections 26—29 gave a rather close fit and permitted to draw — as it seems — significant conclusions about the intercoefficients in question. We shall now discuss an example where the fit is much poorer; its main interest will be to illustrate how the bunch-analysis technique permits us to obtain at least that little regression information which can be squeezed out the data. The example is the potato data for which correlation coefficients are given in Section 2.

The complete set of scatterances in this example was computed. In table (31. 1) are given the lowest and highest values as well as the medians for each dimensionality of the scatterances. (A complete list will be sent on request).

TABLE (31. 1). SCATTERANCES IN THE POTATO DATA.

Set	Median value of scatterance	Lowest scatterance found	Highest scatterance found
2-rowed	.954	.795	.999
3 »	.848	.626	.984
4 »	.708	.536	.893
5 »	.619	.439	.774
6 »	.473	.367	.625
7 »	.369	.309	.425
8 »	.283		

It is quite obvious that the values of the scatterances in the potato data are distinctly lower than for random variates. A glance at (31.1) as compared with (30.3) shows this immediately. Therefore some sort of organisation is undoubtedly present, but, of course, it may not be good enough to permit to point out exactly *how* the various factors influence price. We see for instance that even in the higher sets of the potato data the organisation is very much poorer than in the meat and butter data studied in the previous sections, e. g. the lowest 6-rowed scatterance in the potato data is decidedly higher than the lowest 6-rowed scatterance in meat and butter. Although the scatterances may not be able to discriminate between certain difficult cases of multiple collinearity — as we have seen — yet they do give an unfailable *necessary* criterion: They *must* be small if any good regression equation shall exist. There is therefore no hope of getting a nice looking bunch map for the potato data.

The reason for this poor fit, even in the higher sets of the potato data, must be that one or more variates which are highly important for the determination of the price are not present amongst those here considered. One such variate that suggests itself is the quantity brought to the market: unfortunately exact information about this variate was not available in the present case.

To see if it should not be possible nevertheless to squeeze some information out of the data, a complete bunch map was plotted for all sets up to and including the four-sets.

A first test that was made on this map was to see what happened if the variate No. 1 (price) was added to any of the two-sets. It turned out to be either detrimental or superfluous (according to the criteria of Section 17) in all cases except when added to the set (24) and (25); in these two cases it turned out to be useful. This information we shall now try to verify in other ways. There are four variates involved in the above preliminary conclusion, namely (1245). Let us see what happens if we add one of these variates to any of the two-sets contained in this four-set. Doing this we get the following star map.

TABLE (31.2). STAR MAP FOR THE THREE-SETS IN (1245).

Adding Variate No.	To the two-sets					
	12	14	24	15	25	45
1			*		*	o
2		*		*		o
4	*	.		●	o	
5	*	●	*			

In the first place we see that any of the 2 three-sets (124) or (125) seems promising. Indeed, adding 1 to (24), 2 to (14) or 4 to (12) we get a useful variate (marked with an asterisk in (31.2)). The same is true if we add 1 to (25), 2 to (15) or 5 to (12). But if we add 4 to (15) or 5 to (14) we get an indication of a detrimental variate. Finally, both 1 and 2 are indicated as superfluous in addition to (45). This in connection with the fact that we have the following values of the original correlation coefficients

(31.3)
$$r_{12} = -0.2108$$
$$r_{14} = -0.4526$$
$$r_{24} = +0.0145$$
$$r_{15} = -0.3152$$
$$r_{25} = -0.0547$$
$$r_{45} = +0.4436$$

suggests the following conclusion: The price 1 depends essentially on the two qualities 2 and 4, these two qualities being practically uncorrelated in the material at hand, the set (24) displaying the smallest gross correlation of all the two sets. But on the other hand there is a close relation between the two qualities 4 and 5, so that we may also as an *alternative* consider to express price as a function of the two qualities 2 and 5. *But we must not make an attempt to express price simultaneously as a function* of 4 an 5. The alternative (124) seems to be slightly better than (125).

The above conclusion is corroborated by the following further features of the bunch map. All the bunches in the three-sets that include price indicate that the price depends negatively

on both the two qualities involved. If the bunches in any such three-set should show compatible signs, the bunch exhibiting the intercoefficient between the two qualities ought therefore to indicate a *negative* slope. This is the case in the sets (124) and (125) *but in none of the other three-sets;* in particular the signs are not compatible in (145).

Finally, the star map for the four-set (1245) is

TABLE (31.4). STAR MAP FOR (1245).

Variate added	Intercoefficient between variate Nos.					
	12	14	24	15	25	45
1			○		○	●
2		○		○		○
4	○			●	●	
5	○	○	●			

If the fit in the three sets had been better, there would have been produced more of an "explosion" by passing to the four-sets, but even as it is (31.4) indicates clearly that the set (1245) should not be accepted. In other words, the two three sets (124) and (125) must be kept distinct.

In normalised coordinates the diagonal regression equation in (124) is

$$(31.5) \qquad \xi_1 = -0.892 \ \xi_2 - 0.978 \ \xi_4$$
$$\text{(price)} \qquad \text{(size)} \qquad \text{(colour)}$$

Of the two variates in the right member of (31.5) the above criteria indicate ξ_4 (colour) as the most important.

The concrete knowledge which we have of the data indicates that the result summarised in (31.5) is quite plausible. It is at least a more plausible result than the one which seemed to follow from the ellipsoid method of Section 2.

If we add the significance factor Θ, the regression coefficients of (31.5) can be written

$$(31.6) \qquad -0.892 \cdot/\cdot 0.229 \ \text{and} \ -0.978 \cdot/\cdot 0.460.$$

It is interesting to compare the above final conclusion with the behaviour of the scatterances. The smallest scatterance in the three-sets is that in (145) which is equal to 0.62. If we should let ourselves be guided only by this, we would be led to adopting just that three-set which by the bunch analysis has been characterised as a dangerous inadmissible set. In other words, we just have the second — the dangerous one — of the two alternatives that were discussed in Section 1 in connection with the interpretation of small scatterances. The subscatterances in the set (145) are

(31. 7)   0.80, 0.90 and 0.80

In order to arrive at that interpretation of the scatterances which we now — from the bunch analysis — know is the correct one, we would have to be so conservative as to take the three numbers in (31. 7) as essentially equal. Possibly the whole appearance of the data might have led us to such a scepticism, but at any rate it is evident that the complete bunch analysis furnishes a much more conclusive technique.

32. CONFLUENCE ANALYSIS AS A MEANS OF DETERMINING TREND PARABOLAE AND OTHER CURVE FITTINGS.

A problem that is frequently encountered in time series analysis is to fit a polynomial as a "trend" in a given time series. The question then arises as to how high a degree one should take in the polynomial. The confluence analysis technique may be a help in answering this question. Of course, strictly speaking, a number of terms representing different powers of the time-variate can never be linearly dependent, but it may be that the *approximate* linear connection that exists between such terms over a short interval makes the whole fitting apparatus so much more sensitive to the random disturbances that it would be better to be satisfied with a smaller number of terms. The point where to stop may then be decided by considering the given time series x_t (t denoting time) as well as the successive powers of t as so many variates to be thrown together in a regression analysis and scrutinised by a bunch and star map analysis. In practice it is usually better to use binomial coefficients in t instead of powers.

More generally the same procedure may, of course, be applied if it is wanted to fit to a given series a linear form in any sequence of prescribed functions.

The bunch analysis may also be applied in testing the significance of the information obtained by successive laggings of a given time series. This is equivalent to using the confluence technique in order to determine the significance of the coefficients of a difference equation which the given series is assumed to satisfy. The confluence analysis would tell how high order of the difference equation it would have a meaning to consider.

More generally the same kind of analysis may be applied to test the significance of the information obtained by applying to a given series a number of different linear operations (moving totals with different weight-systems).

33. THE INADEQUACY OF THE CLASSICAL SAMPLING THEORY AS A MEANS OF TESTING LINEAR CONFLUENCY.

In concluding the present investigation it may be interesting to compare the results obtained by bunch analysis with that obtained by the classical sampling error approach in a case of linear confluency.

As an example I select the constructed example in Section 23, where we actually know the composition of the data and thus can see directly which method leads to an approximately correct result and which one gives nonsense results.

The classical sampling error method of testing the "significance" of the coefficients in a regression equation is as follows. If the number of observations is small, the coefficients have to be tested by "Student"'s distribution. But if the number of observations is fairly large — as it is in the constructed example where $N = 100$ — there is only a negligeable difference between the result obtained by using "Student"'s distribution and that obtained by simply computing the ordinary standard errors of the regression coefficients. In the present case it will amply suffice to consider the ordinary standard errors.

The formula for the elementary regression coefficient of the non-normalised variate x_i on x_j taken within the set $(\alpha\beta \dots \gamma)$ is

$$(33.1) \qquad b_{ij(\alpha\beta \dots \gamma)} = -\frac{\hat{m}_{ij(\alpha\beta \dots \gamma)}}{\hat{m}_{ii(\alpha\beta \dots \gamma)}}$$

where the $\hat{m}$'s are the elements of the adjoint moment matrix in the set $(\alpha\beta \dots \gamma)$.

The standard error of (33. 1) according to the classical formulae[1] is

$$(33.2) \qquad \sigma\left[b_{ij(\alpha\beta\ldots\gamma)}\right] = \frac{1}{\sqrt{N'}} \sqrt{\frac{m_{ii}}{m_{jj}}} \cdot \frac{\sqrt{\triangle_{\alpha\beta\ldots\gamma} \cdot \triangle_{\alpha\beta\ldots)ij(\ldots\gamma}}}{\triangle_{\alpha\beta\ldots)i(\ldots\gamma}}$$

where the $\triangle$-s are the scatterances and the $m_{ii} = [x_i x_i]$ the sum squares of the variates extended over all observations. The number N' is the *corrected* number of observations (the degrees of freedom), namely

$$(33.3) \qquad N' = N - \nu$$

where N is the actual number of observations and ν the number of variates in the regression equation, i. e. the number of affixes in the set $(\alpha\beta\ldots\gamma)$.

If the regression equation is written in normal coordinates, the regression coefficient of x_i on x_j in the set $(\alpha\beta\ldots\gamma)$ will be

$$(33.4) \qquad \beta_{ij(\alpha\beta\ldots\gamma)} = \sqrt{\frac{m_{jj}}{m_{ii}}} \, b_{ij(\alpha\beta\ldots\gamma)} = -\frac{\hat{r}_{ij(\alpha\beta\ldots\gamma)}}{\hat{r}_{ii(\alpha\beta\ldots\gamma)}}$$

where the $\hat{r}$-s are the elements of the adjoint correlation matrix in the set $(\alpha\beta\ldots\gamma)$, in other words, they are the elements in the tilling table for the set $(\alpha\beta\ldots\gamma)$. When the tilling technique is used, it is most convenient first to compute the regression equations in normalised coordinates, and then — if wanted — to pass to the coefficients b by means of the first equation in (33.4).

The standard error of (33. 4) is $\sqrt{m_{jj}/m_{ii}}$ times the standard error of (33. 1), that is

---

[1] See for instance *Yule:* An Introduction to Statistics, chapter XVII, or *Ezekiel:* Methods of Correlation Analysis page 258. The standard errors of the regression coefficients are usually given by means of recurrence formulæ to be applied in connection with certain numerical computation schemes. For our purpose it is better to consider the explicite formula (33. 2). This formula is even to be preferred in actual computation whenever the scatterances are available (for instance through the complete tilling).

$$(33.5) \qquad \sigma\left[\beta_{ij;\alpha\beta\,\ldots\,\gamma}\right] = \frac{1}{\sqrt{N'}} \cdot \frac{\sqrt{\triangle_{\alpha\beta\,\ldots\,\gamma}\,\triangle_{\alpha\beta\,\ldots\,ij\,\ldots\,\gamma}}}{\triangle_{\alpha\beta\,\ldots\,i(\ldots\,\gamma}}$$

The formula (11.15) shows that in any subset $(\alpha\beta\ldots\gamma)$ which — so far as the systematic components of the variates is concerned — is multiply collinear, both the scatterance and the subscatterances will have the *disturbances* as their principal terms. In other words, they will be meaningless from the viewpoint of the regression equations studied. In any multicollinear set the standard errors (33.2) and (33.5) will therefore be in point of principle be of the indeterminate $\frac{0}{0}$ form, and in practice they will have to be looked upon as *numbers drawn at random*. In other words, they will be entirely meaningless as tests of the "significance" of the regression coefficients. If we nevertheless use them we actually take a number drawn out of one hat as an expression for the "significance" of some other number drawn out of another hat.

That this is actually so is verified in a striking manner if we compute the elementary regression coefficients and their standard errors in the set (1234) of the constructed example of Section 23. The results are given in (33.6) and (33.7). In these tables the regression coefficients are printed in large types and the corresponding standard errors in small types. The regression coefficients in table (33.6) are obtained simply by reducing the elements on each row in the four-rowed tilling table 3 in Section 23, in such a way that the diagonal element becomes equal to — 1. The other elements in that row will then be the regression coefficients (33.4).

TABLE (33.6). ELEMENTARY REGRESSION COEFFICIENTS AND THEIR STANDARD ERRORS IN THE SET (1234) IN THE CONSTRUCTED EXAMPLE.

	1	2	3	4
1	— 1.0000	— .1126	.7214	.6596
	.0000	.1017	.0692	.0767
2	— .1120	— 1.0000	.7408	— .6593
	.1012	.0000	.0667	.0763
3	.7362	.7599	— 1.0000	.0173
	.0706	.0684	.0000	.1031
4	.6607	— .6637	.0169	— 1.0000
	.0768	.0768	.1012	.0000

TABLE (33. 7). ELEMENTARY REGRESSION COEFFICIENTS AND THEIR
STANDARD ERRORS IN THE SET (12345) IN THE CONSTRUCTED
EXAMPLE.

	1	2	3	4	5
1	— 1.0000	— .1417	.7414	.6423	.0153
	.0000	.1048	.0714	.0780	.0189
2	— .1328	— 1.0000	.7554	— .6341	.0350
	.0982	.0000	.0648	.0745	.0181
3	.7162	.7788	— 1.0000	.0366	— .0343
	.0690	.0668	.0000	.1008	.0133
4	.6472	— .6817	.0382	— 1.0000	.0112
	.0786	.0801	.1046	.0000	.0140
5	.8137	1.9924	— 1.8936	.5932	— 1.0000
	.7482	.7451	.7357	.7424	.0000

If the standard errors should be reliable warning signals,
they ought to tell us to keep away from *any* of these regression
equations. Indeed, from the way in which the example was
constructed we know that not a single one of the regression
coefficients in the tables (33. 6) and (33. 7) has a meaning.

The minimum requirement which the standard errors must
satisfy in order to be such a warning signal is that for any
given regression coefficients the standard error must be equal
to at least one-third or a quarter of the absolute value of the
regression coefficient in question. Otherwise we would conclude
that a least the *sign* of this regression coefficient is significant.
How is this fulfilled in the tables (33. 6) and (33. 7)? *It is very
far from fulfilled.* Take for instance the first equation in (33. 6).
Here we have $\beta_{13} = 0.7214$ and its standard error 0.07. In other
words, the standard error is less the one-tenth of the regression
coefficient. And for the next regression coefficient β_{14} the
standard error is about one-tenth. No statistician who is used
to working with standard errors would hesitate to conclude that
the last two regression coefficients are significant. At least he
would conclude that it is practically certain that both these
coefficients are *positive*. From the way in which the example
was constructed we know that this is sheer nonsense; a re-
gression equation in the set (1234) has indeed no meaning at all..

In the second equation of (33. 6) we have a similar situation.
One would here conclude for instance that β_{23} is significantly
positive and β_{24} significantly negative, and so on.

In view of the fact that the standard errors are now to be

looked upon as numbers drawn at random, we would expect that, roughly speaking, one-half of them will actually do the right thing, namely to warn us against the regression coefficient in question, and the other half would do the wrong thing, namely *not* to warn us. Disregarding the diagonal elements that are by necessity equal to — 1, we have in (33. 6) 12 standard errors of which 8 do the wrong thing. In (33. 7) we may disregard the coefficients in the last column where — as we have previously seen — a persistency effect is present. This leaves us with 16 coefficients of which 8 do the wrong thing.

The bunch analysis of Section 24, it will be remembered, furnished the correct criteria for the nonsense of the regression equation in the sets (1234) and (12345), and did it with such distinctness that there could be no doubt about the conclusiveness of the result.

<div align="center">*</div>

I do not claim that the technique developed in the present paper will, like a stone of the wise, solve all the problems of testing "significance" with which the economic statistician is confronted. No statistical technique, however refined, will ever be able to do such a thing. The ultimate test of significance must consist in a netword of conclusions and cross checks where theoretical economic considerations, intimate and realistic knowledge of the data and a refined statistical technique concur. But I do claim that the technique here presented will in a great number of cases be very helpful. I would even venture to say that for many kinds of problems it will be *indispensable* — until something better is found that can replace it.